

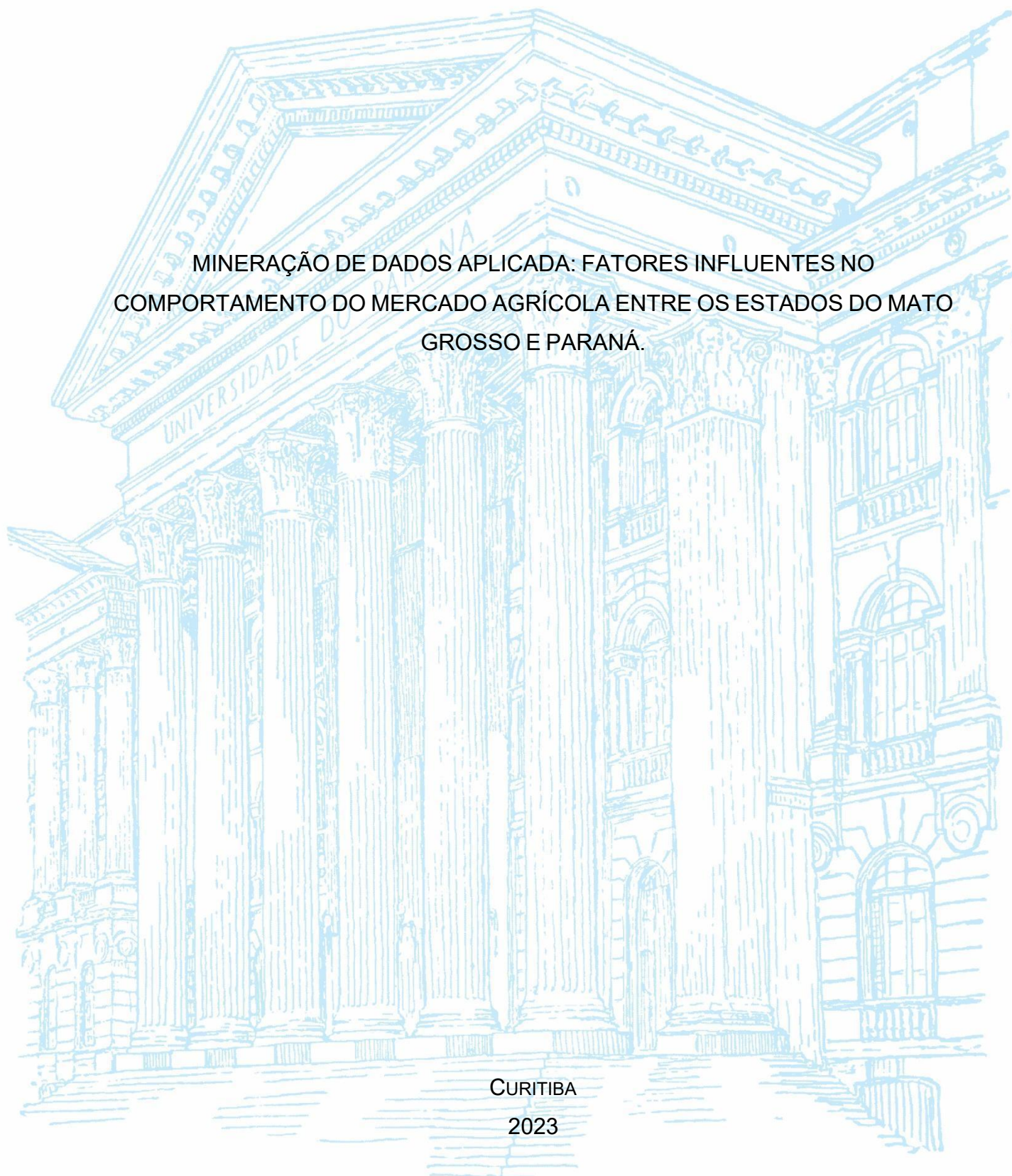
UNIVERSIDADE FEDERAL DO PARANÁ

BARBARA DA COSTA KLECHOVICZ

MINERAÇÃO DE DADOS APLICADA: FATORES INFLUENTES NO
COMPORTAMENTO DO MERCADO AGRÍCOLA ENTRE OS ESTADOS DO MATO
GROSSO E PARANÁ.

CURITIBA

2023



BARBARA DA COSTA KLECHOVICZ

MINERAÇÃO DE DADOS APLICADA: FATORES INFLUENTES NO
COMPORTAMENTO DO MERCADO AGRÍCOLA ENTRE OS ESTADOS DO MATO
GROSSO E PARANÁ.

Trabalho de Conclusão de Curso apresentado ao curso de Gestão da Informação, do Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Gestão da Informação

Orientador: Prof. Dr. José Marcelo Almeida Prado Cestari.

CURITIBA

2023

Descobrir consiste em olhar para o que todo mundo está vendo e pensar
uma coisa diferente (Roger Von Oech)

AGRADECIMENTOS

Primeiramente, agradeço a Deus por Sua constante orientação e força, que me sustentaram durante todo o percurso acadêmico.

À minha família, verdadeira base de apoio, agradeço por cada palavra de incentivo, paciência e compreensão nos momentos desafiadores.

Ao meu amigo Henrique Bernini, a gratidão é imensurável. Sua amizade, conselhos e presença constante foram combustíveis essenciais para superar obstáculos dessa jornada acadêmica.

Ao meu orientador, José Cestari, expresso minha profunda admiração e reconhecimento. Sua orientação expert e dedicação foram cruciais para o desenvolvimento deste trabalho.

A todos os amigos, colegas e professores que de alguma forma contribuíram para este percurso, meu sincero agradecimento. Cada interação e troca de conhecimento deixou uma marca valiosa em minha jornada acadêmica.

Muito obrigado a todos!

Resumo

A pesquisa aborda a análise do mercado agrícola nos estados do Mato Grosso e Paraná por meio da mineração de dados. A pesquisa, conduzida mediante a aplicação de técnicas de mineração de dados em diversas bases governamentais, almeja identificar e compreender os fatores que impactam esse setor. Através da análise, o estudo revela correlações significativas, evidenciando a eficácia da mineração de dados na investigação das dinâmicas desses mercados. A conclusão mostra a peculiaridade de cada estado em relação ao que influencia seu comportamento no mercado agrícola. Assim, promove-se uma análise mais ampla e informada do comportamento desses mercados, fundamentais para a economia brasileira.

Palavras-chave: Mineração de dados; mercado agrícola; Gestão da informação

Abstract

This research addresses the analysis of the agricultural market in the states of Mato Grosso and Paraná through data mining. Conducted by applying data mining techniques to various government databases, the study aims to identify and understand the factors impacting this sector. Through analysis, the research reveals significant correlations, showcasing the effectiveness of data mining in investigating the dynamics of these markets. The conclusion highlights the uniqueness of each state concerning the factors influencing its behavior in the agricultural market. This promotes a broader and informed analysis of the behavior of these markets, which are fundamental to the Brazilian economy.

Keywords: Data mining; agricultural market; Information management

LISTA DE ILUSTRAÇÕES

FIGURA 1: GERENCIAMENTO DE INFORMAÇÕES	19
FIGURA 2: ESTÁGIOS DO KDD	20
FIGURA 3: CONEXÕES TRANSDISCIPLINARES DA MINERAÇÃO DE DADOS.....	22
FIGURA 4: ETAPA DO PROCESSO KDD E PRINCIPAIS TAREFAS E TÉCNICAS RELACIONADAS.....	23
FIGURA 5: RANKING DA PRODUÇÃO AGRÍCOLA DO BRASIL (2021)	27
FIGURA 6: RANKING DA PRODUÇÃO AGRÍCOLA DO PARANÁ (2021).....	27
FIGURA 7: RANKING DA PRODUÇÃO AGRÍCOLA DO MATO GROSSO (2021).....	28
FIGURA 8: CARACTERIZAÇÃO DA PESQUISA.....	34

LISTA DE GRÁFICOS

GRÁFICO 1: CORRELAÇÃO DOS ATRIBUTOS NO MATO GROSSO E NO MILHO.....	38
GRÁFICO 2: CORRELAÇÃO DOS ATRIBUTOS NO PARANÁ E NO MILHO.....	38
GRÁFICO 3: CORRELAÇÃO DOS ATRIBUTOS NO MATO GROSSO E NA SOJA	39
GRÁFICO 4: CORELAÇÃO DOS ATRIBUTOS NO PARANÁ E NA SOJA	39
GRÁFICO 5: HISTORICO DE TREINAMENTO MATO GROSSO MILHO	45
GRÁFICO 6: HISTORICO DE TREINAMENTO MATO GROSSO SOJA.....	46
GRÁFICO 7: HISTORICO DE TREINAMENTO PARANÁ MILHO	47
GRÁFICO 8: HISTORICO DE TREINAMENTO PARANÁ SOJA.....	47

LISTA DE QUADROS

QUADRO 1: DISPARIDADE ENTRE DADO, INFORMAÇÃO E CONHECIMENTO	18
QUADRO 2: ETAPA DO PROCESSO KDD E PRINCIPAIS TAREFAS E TÉCNICAS RELACIONADAS	24
QUADRO 3: HEURÍSTICAS DA TÉCNICA PREDITIVA.....	24
QUADRO 4: NCM	32
QUADRO 5: ATRIBUTOS SELECIONADOS.....	34
QUADRO 6: ATRIBUTOS CORRIGIDOS	35
QUADRO 7 : TRANSFORMAÇÃO DE DADOS	37
QUADRO 8: DESCRIÇÃO ATRIBUTOS	37
QUADRO 9: SIGNIFICADO DOS PESOS	41

Sumário

1.	INTRODUÇÃO	16
1.1	JUSTIFICATIVA E PROBLEMATIZAÇÃO	16
1.2	OBJETIVOS	17
1.2.1	Objetivo geral	17
1.2.2	Objetivos específicos.....	17
2	REVISÃO DE LITERATURA.....	18
2.1	GESTÃO DA INFORMAÇÃO	18
2.2	KDD	19
2.3	MINERAÇÃO DE DADOS	21
2.4	MERCADO AGRÍCOLA.....	25
3	METODOLOGIA.....	29
3.1	CARACTERIZAÇÃO DA PESQUISA	30
3.2	AMBIENTE DE DESENVOLVIMENTO.....	30
3.3	MATERIAIS E MÉTODOS.....	31
3.3.1	Base de dados.....	31
3.3.2	Ferramentas	33
3.4	PROCEDIMENTOS METODOLÓGICOS	33
4	PROCESSAMENTO E ANÁLISE DOS RESULTADOS	34
4.2	PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS	35
4.3	TRANSFORMAÇÃO DOS DADOS	37
4.4	ESTATÍSTICA DESCRITIVA.....	37
5.	CONCLUSÃO.....	48
6	REFERÊNCIAS.....	49
5	APÊNDICE	53
5.1	APÊNDICE A - ATRIBUTOS DESCARTADOS	53

1. INTRODUÇÃO

A mineração de dados desempenha um papel crucial na era da informação, permitindo explorar grandes quantidades de dados de forma eficiente e identificar conhecimentos valiosos que seriam difíceis de serem descobertos por meio de métodos tradicionais. Silva (2016), define a mineração como, exploração analítica automatizada ou semiautomática de grandes bases de dados visa descobrir padrões relevantes para embasar a assimilação de informações importantes e gerar conhecimento. Ela pode ser utilizada em diversas áreas, um exemplo é no mercado agrícola.

A agricultura brasileira é fornecedora de empregos, riquezas, alimentos e bioenergia para o Brasil e outros países. É um dos setores que mais contribui para o crescimento do PIB nacional e que responde por 21% da soma de todas as riquezas produzidas, um quinto de todos os empregos (Embrapa, 2020).

Além de contribuir internamente com o país, a agricultura brasileira fornece alimento e insumo para diversos países. Segundo dados do Ministério do Desenvolvimento, Indústria, Comércio e Serviços (2023), até a 2ª Semana de Junho/2023, houve crescimento de 25,4% em Agropecuária, que somou US\$ 2,77 bilhões no mercado de exportação. Um nicho importante deste setor é o de *commodities*. Martins (2009) refere-se ao *comodities* como mercadorias em estado bruto ou de simples industrialização, que podem ser estocadas e produzidas em grandes quantidades.

Esse mercado gera dados constantemente e a abundância destes dados podem contribuir para tomada de decisão. Ferraz (2022) aborda a mineração de dado como processo que permite gerar conhecimento através de extração em base de dados.

Nesse sentido, o objetivo dessa pesquisa é analisar bases de dados referente ao que movimenta o mercado agrícola para responder à pergunta. **Como a mineração de dados pode ajudar na comparação do comportamento dos mercados de soja e milho nos estados do Paraná e Mato Grosso?**

1.1 JUSTIFICATIVA E PROBLEMATIZAÇÃO

Considerando o atual panorama do mercado agrícola, a justificativa para a realização deste estudo surge do interesse empreendido na área da agricultura. Vale ressaltar que, apesar da vital importância da agricultura, há uma carência de trabalhos

acadêmicos específicos sobre este tema dentro do curso de Gestão da Informação dentro do acervo digital consta apenas dois trabalhos relacionado ao tema. Esta lacuna representa uma oportunidade relevante para a contribuição com conhecimento científico original.

Dessa forma, o presente trabalho partiu da necessidade de entender como o estudo no mercado agrícola pode impactar direta ou indiretamente agricultores, empresas de médio e grande porte, inclusive órgãos governamentais. Os resultados desta pesquisa contribuirão para o entendimento dos desafios e oportunidades enfrentados pelos agricultores e agentes do setor, e poderão auxiliar no desenvolvimento de estratégias mais eficazes para impulsionar o crescimento sustentável do mercado agrícola

1.2 OBJETIVOS

Com o propósito de se obter respostas para o problema de pesquisa, foram definidos o objetivo geral e específicos;

1.2.1 Objetivo geral

Analisar o comportamento do mercado agrícola nos estados do Mato Grosso e Paraná por meio da mineração de dados, destacando os fatores que mais impactam esse mercado, com foco na relação entre o valor do produto e os preços do diesel ao longo do tempo.

1.2.2 Objetivos específicos

- i. Identificar, otimizar e preparar a base para a realização da mineração;
- ii. Relacionar as bases de dados e selecionar os atributos a serem analisados;
- iii. Utilizar as estatísticas descritivas e definir os algoritmos de mineração de dados para análise da base;
- iv. Aplicar os algoritmos escolhidos;
- v. Comparar os resultados.

2 REVISÃO DE LITERATURA

Nesta seção, iremos fornecer um breve embasamento teórico com o objetivo de fundamentar o estudo em questão. Para facilitar o entendimento ficará dividido em Gestão da Informação; Knowledge Discovery in Databases (KDD); mineração de dados; mercado agrícola.

2.1 GESTÃO DA INFORMAÇÃO

Gestão da informação refere-se ao processo de coleta, organização, armazenamento, análise, disseminação e utilização estratégica de informações dentro de uma organização. Segundo Silva et. al. (2021), a gestão da informação abrange uma série de iniciativas que visam atender às necessidades informacionais, mapear os fluxos formais de informação e realizar atividades como coleta, filtragem, análise, organização, armazenamento e disseminação, com o propósito de apoiar o desenvolvimento de atividades e tomada de decisões no contexto organizacional.

Dentro da gestão da informação é imprescindível o domínio dos conceitos, dados, informação e conhecimento. Davenport (1998, p.18) apresentada através de uma abordagem definindo inicialmente os conceitos conforme ilustra o quadro 1.

Quadro 1: DISPARIDADE ENTRE DADO, INFORMAÇÃO E CONHECIMENTO

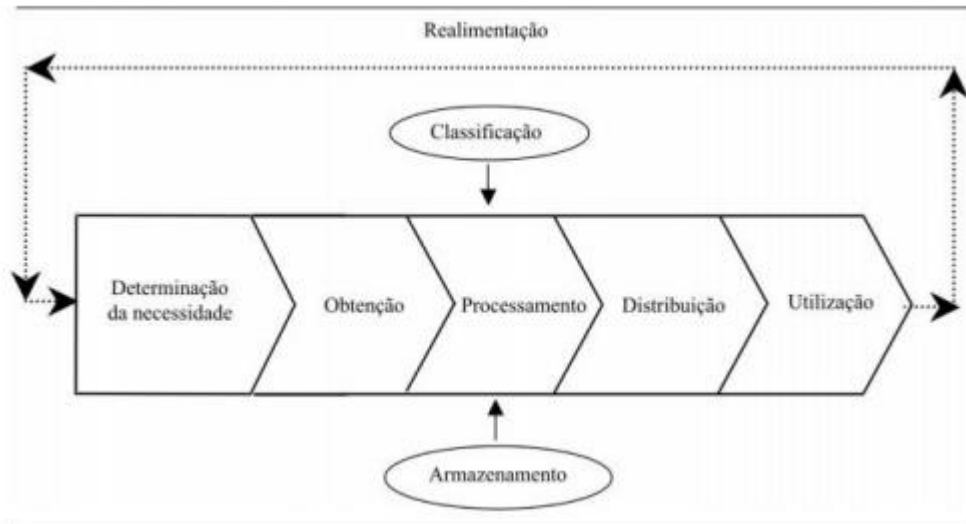
Dados	Informação	Conhecimento
<p>Simple observações sobre o estado do mundo.</p> <ul style="list-style-type: none"> ● Facilmente estruturados; ● Facilmente obtido por máquinas; ● Frequentemente quantificado; ● Facilmente transferível; 	<p>Dados dotados de relevância e propósito.</p> <ul style="list-style-type: none"> ● Requer unidade de análise; ● Exige consenso em relação ao significado; ● Exige necessariamente a mediação humana; 	<p>Informação valiosa da mente humana.</p> <ul style="list-style-type: none"> ● De difícil estruturação; ● De difícil captura em Máquinas; ● Frequentemente tácito; ● De difícil transferência;

FONTE: adaptado de Davenport (1998, p. 18).

Nas organizações as informações contêm fluxos e para Beal (2008, p. 29) é descrito como: “identificação de necessidades e requisitos; obtenção; tratamento; armazenamento; distribuição; uso; descarte”.

Moraes e Filho (2006, p.125) apresenta divide em cinco etapas o processo de gerenciamento de informações, conforme apresentado na Figura 1.

Figura 1: GERENCIAMENTO DE INFORMAÇÕES



FONTE: Moraes; Filho (2006, p.126)

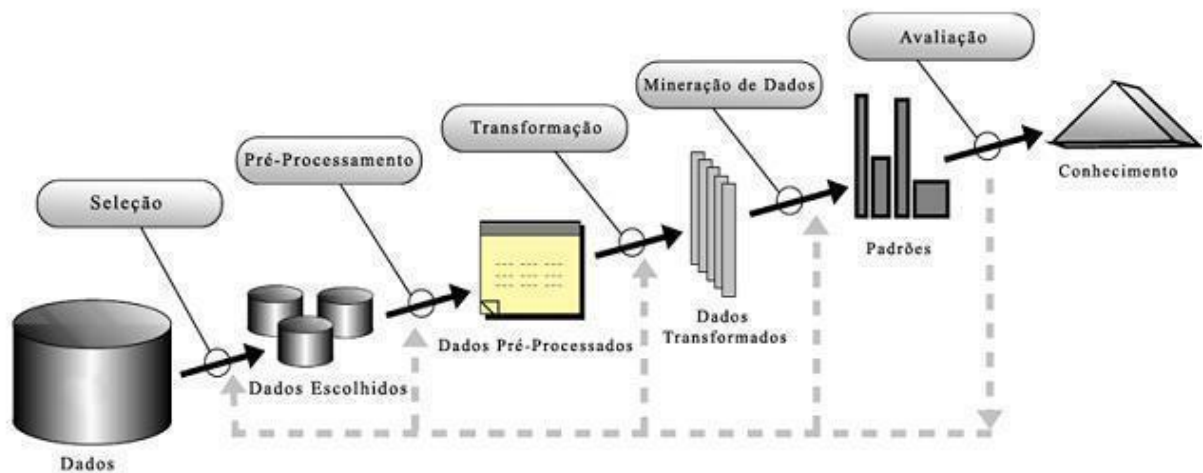
De acordo com as informações, segundo Moraes e Filho (2006, p.125) A primeira consiste na determinação das necessidades de informação dos clientes, identificando as fontes de informação necessárias para atender às principais demandas. Em seguida, temos a etapa de obtenção, na qual ocorre a coleta de dados que serão utilizados nas demais atividades. A terceira parte do fluxo é o processamento, uma fase crucial que define a melhor forma de acessar as informações e determina como elas devem ser armazenadas. Na sequência, temos a distribuição e apresentação da informação, onde são estabelecidas as metodologias mais adequadas para apresentar e disponibilizar os dados. Por fim, chegamos à etapa de utilização, na qual os clientes consomem os resultados no processo estratégico e nas tomadas de decisão.

2.2 KDD

A descoberta de conhecimento em bases de dados, conhecida como Knowledge Discovery in Databases (KDD), é um processo complexo que envolve a identificação de padrões em grandes conjuntos de dados. O objetivo do KDD é encontrar padrões intrínsecos aos dados e apresentá-los de forma a facilitar sua assimilação como conhecimento (DA SILVA; PERES; BOSCARIOLI, 2016).

O processo de KDD é composto por várias etapas. Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), existem cinco etapas principais: seleção dos dados, pré-processamento e limpeza dos dados, transformação dos dados, data mining e interpretação/avaliação dos resultados. Estas fases estão ilustradas na Figura 2.

Figura 2: ESTÁGIOS DO KDD



FONTE: Fayyad, Piatetsky-Shapiro e Smyth (1996)

A primeira etapa, seleção dos dados, envolve a escolha do conjunto de dados ou de um subconjunto específico de variáveis ou amostras de dados para análise. Em seguida, na etapa de pré-processamento e limpeza dos dados, são realizadas atividades como remoção de ruído e dados inconsistentes, integração de dados de múltiplas fontes, seleção ou redução de dados relevantes e transformação dos dados em formatos apropriados para a mineração.

A etapa de transformação dos dados é responsável por encontrar recursos úteis para representar os dados, dependendo dos objetivos da análise. Isso pode envolver a discretização de atributos contínuos, eliminação de atributos redundantes ou irrelevantes, entre outras transformações.

Em seguida, na etapa de data mining, são aplicados algoritmos de mineração de dados para extrair conhecimento dos dados pré-processados. Esses algoritmos podem ser usados para realizar análises descritivas, como medidas de distribuição e visualização, agrupamento, predição e detecção de anomalias.

Já na etapa de interpretação/avaliação dos resultados, os padrões descobertos são avaliados e interpretados com o objetivo de identificar conhecimentos úteis e não triviais. Essa etapa requer conhecimento de estatística e das técnicas utilizadas no processo de data mining

A descoberta de conhecimento em bases de dados é um campo de estudo que abrange diversas disciplinas e combina diferentes áreas de conhecimento. Há destaque nas áreas como estatística, descoberta de padrões, banco de dados, inteligência computacional e aprendizado de máquina (Goldshmidt & Passos, 2005).

A aplicação adequada das etapas do KDD é fundamental para obter resultados corretos e úteis. Negligenciar alguma etapa ou não adequar os dados corretamente pode levar a respostas errôneas para problemas corretos (Sharda, Delen & Turban, 2019).

No processo de data mining, são aplicados algoritmos específicos que podem variar dependendo das tarefas de mineração de dados. Essas tarefas incluem classificação, regressão, agrupamento e associação, cada uma com seus respectivos objetivos e técnicas

Dantas et al., 2008 ressalta que a etapa de interpretação dos resultados é crucial para a tomada de decisão. A interpretação dos padrões descobertos requer habilidades de análise e experiência para associá-los de maneira adequada e utilizar o conhecimento extraído para a tomada de decisões estratégicas.

2.3 MINERAÇÃO DE DADOS

A mineração de dados é uma fase essencial do processo de descoberta de conhecimento em bases de dados (KDD), que consiste em quatro partes principais: base de dados, preparação de dados, mineração de dados e avaliação do conhecimento (Castro e Ferrari, 2016). Os autores também discutem que o processo de mineração de dados tem como objetivo extrair conhecimento, o qual pode ser aplicado na tomada de decisões estratégicas em diversas áreas. A mineração de dados é uma disciplina interdisciplinar e multidisciplinar que abrange conhecimentos de várias áreas, como banco de dados, estatística, aprendizado de máquina, reconhecimento de padrões e inteligência artificial. (Castro e Ferrari, 2016). Como ilustrado na Figura 3.

Figura 3: CONEXÕES TRANSDISCIPLINARES DA MINERAÇÃO DE DADOS



FONTE: Castro e Ferrari (2016).

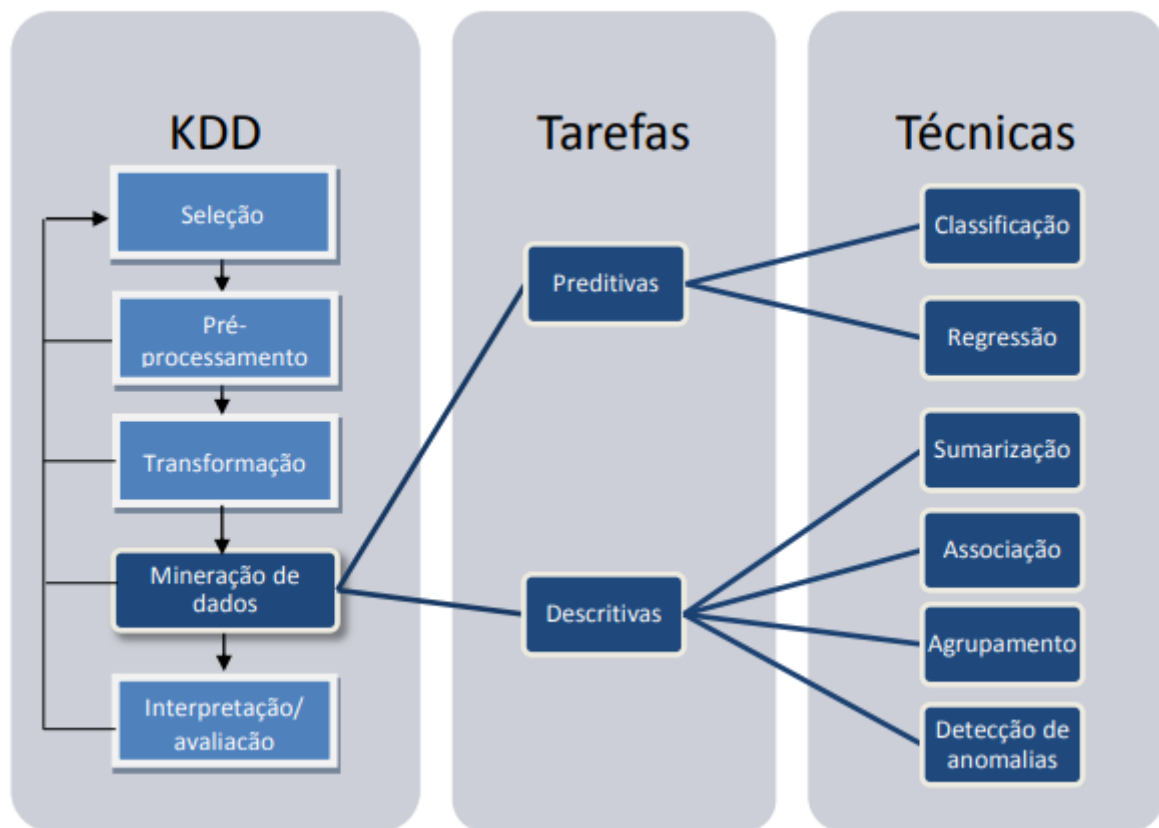
Os dados utilizados na mineração podem ser de diferentes tipos, classificados como atributos qualitativos - nominais e ordinais - ou atributos quantitativos - intervalares e proporcionais (Tan, Steinbach e Kumar, 2009). É fundamental realizar uma preparação prévia dos dados, incluindo limpeza, integração, seleção e transformação, para garantir que a análise seja eficiente e eficaz (Castro e Ferrari, 2016). Os autores ainda advertem que a preparação de dados é uma etapa essencial para a obtenção de resultados precisos e relevantes durante o processo de mineração de dados.

A mineração de dados envolve tarefas descritivas e preditivas. Segundo Castro e Ferrari (2016), as tarefas descritivas, são as que buscam caracterizar as propriedades gerais dos dados, e preditivas, são as que realizam inferências a partir dos dados com o objetivo de fazer previsões. As tarefas descritivas da mineração de dados são muitas vezes exploratórias em sua natureza e frequentemente requerem

técnicas de pós-processamento para validar e explicar os resultados. (TAN; STEINBACH; KUMAR, 2009, p. 8)

Expandindo essa perspectiva, Nogueira (2015) ressalta a relevância da mineração de dados como um componente crucial no processo de descoberta de conhecimento. O autor destaca as diversas tarefas envolvidas nesse processo e apresenta as técnicas mais frequentemente utilizadas para a implementação dessas tarefas, como ilustrado na Figura 4.

Figura 4: ETAPA DO PROCESSO KDD E PRINCIPAIS TAREFAS E TÉCNICAS RELACIONADAS



FONTE: Nogueira (2015)

As tarefas descritivas da mineração de dados possuem uma natureza exploratória e, frequentemente, exigem a aplicação de técnicas de pós-processamento para validar e elucidar os resultados obtidos (Tan, Steinbach e Kumar, 2009).

Souza (2021) complementa que de maneira geral, as tarefas descritivas da mineração de dados têm como objetivo medir, explorar e descrever as características intrínsecas dos dados, permitindo a investigação das distribuições de frequência, medidas de centralidade, variabilidade, posição relativa e associação dos dados. A

autora apresenta de forma sintetizada heurísticas das tarefas descritivas como mostra no Quadro 2 abaixo.

Quadro 2: ETAPA DO PROCESSO KDD E PRINCIPAIS TAREFAS E TÉCNICAS RELACIONADAS

Heurísticas	Descrição
Agrupamento	As técnicas de agrupamento têm a finalidade de identificar grupos de instâncias a partir de seus atributos. Estes grupos são conhecidos como Clusters e eles maximizam as similaridades do dado dentro do grupo.
Associação	As técnicas de associação têm como finalidade encontrar relações ou dependências entre as instâncias de um atributo da análise do conjunto de dados.
Deteção de Anomalias	As técnicas de deteção de anomalias ou também conhecida como deteção de outliers, ajuda na identificação de anomalias nos dados, removendo dados discrepantes que não enquadram no comportamento dos demais registros.
Sumarização	As técnicas de Sumarização são a generalização dos dados, ela extrai informações condensadas de uma massa de volume de dados e fornece um resultado com um conjunto menor, tendo uma visão mais geral dos dados.

FONTE: Souza (2021)

As tarefas preditivas podem ser realizadas utilizando técnicas de classificação ou regressão, enquanto as tarefas descritivas abrangem técnicas de sumarização, associação, agrupamento e deteção de anomalias. (Tan, Steinbach e Kumar, 2009). Souza (2021) apresenta de forma sintetizada heurísticas das tarefas descritivas como mostra abaixo no Quadro 3.

Quadro 3: HEURÍSTICAS DA TÉCNICA PREDITIVA

Heurísticas	Descrição
Classificação	Esta técnica busca encontrar modelos de predição para variáveis meta do tipo contínuas e ela utiliza todos os atributos que compõem para predisse os resultados.

Regressão(estimação)	Já a técnica de regressão também prediz o resultado e ela tem como classe numérica.
-----------------------------	---

FONTE: Souza (2021)

O processo de mineração de dados envolve várias etapas, e o pós-processamento desempenha um papel importante na análise e interpretação dos resultados obtidos. Segundo Milani e Carvalho (2013), o pós-processamento tem como objetivo eliminar a redundância, generalizar e identificar no conjunto os padrões mais relevantes para a solução do problema em questão.

Uma das atividades realizadas durante o pós-processamento é a visualização da análise, conforme mencionado por Boente, Oliveira e Rosa (s.d., p. 9). A visualização permite uma representação gráfica dos resultados, facilitando a compreensão e a interpretação dos padrões identificados.

A capacidade de extrair conhecimentos úteis e significativos a partir dos dados torna a mineração de dados uma ferramenta valiosa para a tomada de decisões em várias áreas.

2.4 MERCADO AGRÍCOLA

A importância da agricultura para o crescimento de um país sempre foi um tópico de grande relevância dentro da teoria econômica (Figueredo, Barros e Guilhoto, 2005)

De acordo com Moura et al. (1999), estudos realizados indicam que o setor agrícola, quando estimulado, pode ter um efeito benéfico sobre o restante da economia, gerando efeitos indutores positivos para outros setores. Além disso, eles demonstraram que ao priorizar as atividades nesse setor como impulsionadoras do crescimento econômico, um país fará uma escolha que promove uma alocação mais eficiente de seus recursos, especialmente no caso de países que ainda não atingiram os padrões de desenvolvimento considerados satisfatórios.

A agricultura apresenta características distintas de outros setores da economia, como maior dependência das condições climáticas, sazonalidade da oferta e instabilidade de preços e renda para os produtores (Buainain; Filho, 2007). Isso gera maiores riscos e incertezas, o que justifica a necessidade de políticas agrícolas que

considerem as especificidades do setor e garantam uma adequada oferta de produtos (Buainain; Filho, apud VEIGA, 2007).

Nesse sentido, o governo desempenha um papel importante no mercado agrícola por meio de políticas agrícolas, como a definição de preços mínimos de venda e o fornecimento de crédito para os produtores. No entanto, a eficácia dessas políticas pode variar, e é necessário considerar as peculiaridades do mercado agrícola e as flutuações nos preços e custos de produção (Nantes, Scaerpelli e Filho, 2008).

O Brasil, como o maior país do continente sul-americano e o quinto maior do mundo, possui uma área de 851 milhões de hectares. Além disso, suas condições climáticas favoráveis permitem o cultivo de uma variedade de produtos agrícolas, tanto de clima temperado quanto tropical. Dourado (2005) salienta que por causa dessa vasta extensão territorial, o Brasil possui um potencial para expandir ainda mais suas áreas cultiváveis, sendo considerado o país com maior capacidade nesse aspecto em todo o mundo.

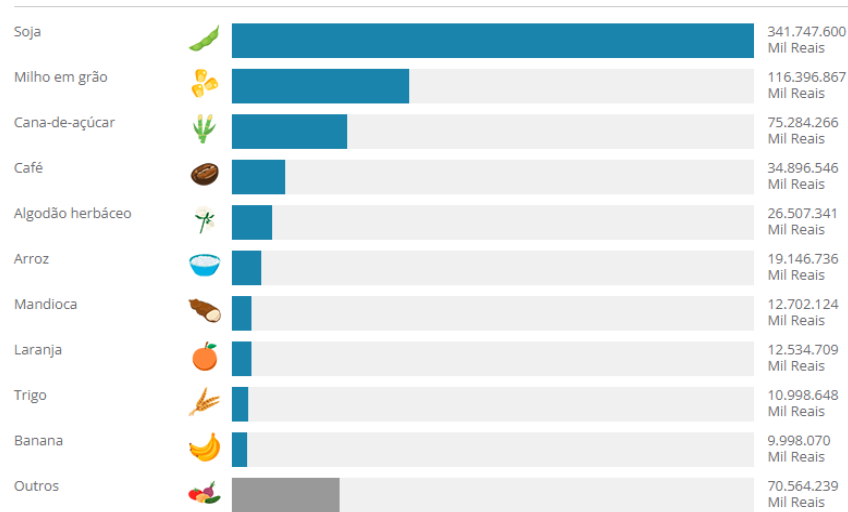
Dourado (2005) traz comparações importantes sobre potencial de aplicação:

Se compararmos com os EUA, se verifica que 19% do território americano é de área plantada (174 milhões de hectares) e 22% (199 milhões de hectares) é ocupado com pastagem, enquanto o Brasil apresenta cultivo de lavouras anuais e permanentes, ocupando apenas 6,7% do território nacional (57 milhões de hectares), enquanto 25,8% são ocupados com pastagens (cerca de 220 milhões de hectares)

A autora ainda expõe que o Brasil tem alcançado consecutivos recordes na produção de grãos, graças ao aumento das médias de produtividade. No entanto, o crescimento das áreas plantadas tem sido proporcionalmente menor em relação a esse incremento.

O Brasil possui produtos agrícolas distribuídos pelas cinco regiões existentes: Norte, Nordeste, Sul, Sudeste e Centro-Oeste. Cada região se especializa em determinados cultivos, levando em consideração as características específicas de clima e solo (Roncon, 2011). A Figura 5, ilustra os principais produtos cultivados no país.

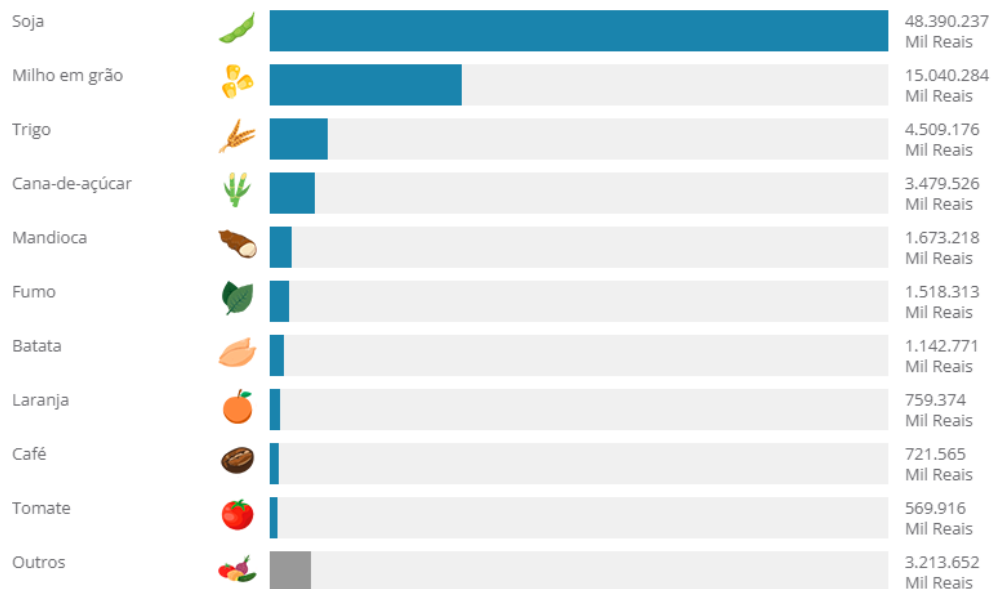
Figura 5: RANKING DA PRODUÇÃO AGRÍCOLA DO BRASIL (2021)



Fonte: IBGE (2023)

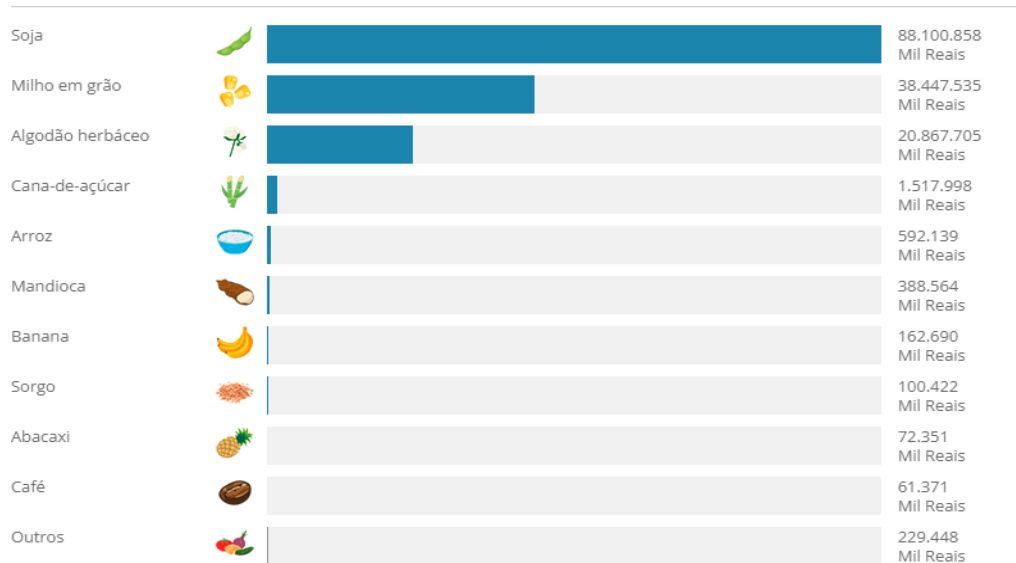
A imagem evidencia que o mercado de *commodities* em especial a soja e milho. A figura 6 e 7 mostra o ranking da produção dos estados do Paraná e Mato Grosso respectivamente.

Figura 6: RANKING DA PRODUÇÃO AGRÍCOLA DO PARANÁ (2021)



Fonte: IBGE (2023)

Figura 7: RANKING DA PRODUÇÃO AGRÍCOLA DO MATO GROSSO (2021)



Fonte: IBGE (2023)

É possível observar a predominância das culturas de soja e milho tanto no estado do Paraná quanto no Mato Grosso, alinhando-se com o cenário agrícola nacional. O Brasil, por sua vez, destaca-se como o maior produtor mundial de soja, sendo responsável por aproximadamente 30% da produção global desse grão (EMBRAPA, 2022).

De acordo com Dall'Agnol (2000), a soja desempenhou um papel fundamental no surgimento da agricultura comercial no Brasil. Ela impulsionou a mecanização das lavouras, modernizou o transporte e expandiu a fronteira agrícola. Além disso, a cultura da soja também teve um efeito positivo no desenvolvimento de outras culturas agrícolas e foi um importante fator de incentivo para o crescimento da avicultura e suinocultura no país. A geração de tecnologias também contribuiu para o aumento da produção de soja no Brasil, elevando-o à posição de segundo maior produtor mundial desse grão.

Silva, Lima e Batista (2011) explicam que a importância do complexo de soja para o Brasil é evidenciada tanto pelo incrível crescimento na produção dessa leguminosa quanto pelas receitas geradas com as exportações de soja em grão e seus derivados, como óleo e farelo de soja. Essa cadeia produtiva abrangente vai além, pois animais alimentados com rações à base de farelo de soja geram

subprodutos que impulsionam outras áreas da economia, como o setor de couro, fertilizantes orgânicos e outros segmentos relacionados (SANCHES, MICHELLON, ROESSING, 2005).

O milho por sua vez, é a cultura agrícola mais produzida no mundo (SYNGENTA, 2022). Duarte, Mattoso e Garcia (2021) expõem que o milho possui uma importância econômica significativa devido às suas diversas formas de utilização, que vão desde a alimentação animal até a indústria de alta tecnologia. Os autores ainda afirmam que de fato, o consumo de milho em grão na alimentação animal representa a maior parte do consumo mundial desse cereal, totalizando cerca de 70%. Embora o consumo humano de milho em grão e seus derivados não tenha uma participação tão expressiva, em regiões de baixa renda, ele desempenha um papel importante como fonte de alimentação, sendo, em algumas situações, a base da dieta diária.

No Brasil, assim como em outros países ao redor do mundo, o milho desempenha um papel fundamental como insumo na produção de ração animal. Souza, Silveira e Ballini (2021) relatam que conforme a criação de aves, suínos, gado para corte e gado leiteiro aumenta, a demanda por esse cereal também cresce, o que implica em uma necessidade de maior produção nacional para suprir essa demanda. Os autores ainda esclarecem:

Neste sentido, o Brasil possui condições de atender sua crescente demanda graças a duas safras anuais: a primeira safra (de verão) e a segunda safra (semeada no período entre março e maio, especialmente no Paraná, São Paulo, Minas Gerais e Centro-Oeste, era conhecida antigamente como safrinha, por ser complementar à safra de verão em volume, porém, nos últimos 20 anos a mesma vê sua importância se inverter, com seu volume passando a ser quase quatro vezes o volume produzido na safra de verão

Segundo a Syngenta (2022) a produção de milho no Brasil cresceu 75% nos últimos 22 anos.

3 METODOLOGIA

Esta seção apresenta a caracterização da pesquisa, o seu ambiente de desenvolvimento.

3.1 CARACTERIZAÇÃO DA PESQUISA

Nesta pesquisa, optou-se por uma abordagem aplicada, visando responder a questões específicas por meio de uma perspectiva descritiva. A escolha dessa abordagem permite a coleta e comparação de diversas informações ao longo do estudo, visando descrever as características de um determinado custo ou fenômeno e estabelecer relações entre variáveis.

Quanto à natureza, esta pesquisa será conduzida de forma predominantemente descritiva. Isso implica a análise minuciosa e sistemática de características do fenômeno estudado, proporcionando uma visão aprofundada de seus aspectos relevantes. A descrição pormenorizada busca fornecer uma compreensão clara das variáveis envolvidas, contribuindo para a formulação de conclusões mais precisas.

A abordagem adotada nesta pesquisa é predominantemente quantitativa, pois se busca realizar análises por meio de ferramentas estatísticas. A ênfase na abordagem quantitativa permite uma avaliação objetiva dos dados, proporcionando resultados mais precisos e embasados.

No estágio inicial desta pesquisa, será realizada uma análise exploratória para mapear e compreender de forma mais ampla o cenário em questão. A exploração inicial será fundamental para identificar variáveis-chave e estabelecer direcionamentos para as fases subsequentes da pesquisa.

Os procedimentos técnicos adotados nesta pesquisa incluirão a aplicação de análises estatísticas, aproveitando ferramentas específicas para avaliação quantitativa. Métodos estatísticos serão empregados de maneira sistemática, proporcionando interpretação dos dados coletados.

3.2 AMBIENTE DE DESENVOLVIMENTO

O ambiente de pesquisa estabelece o cenário para a análise abrangente do comportamento do mercado agrícola entre os estados do Paraná e Mato Grosso. Este estudo busca compreender os fatores que influenciam as decisões dos produtores agrícolas quanto à exportação, uma análise essencial dada a importância econômica

e agrícola desses estados no contexto nacional.

Para fundamentar nossa análise, utilizaremos séries históricas de dados abertos disponibilizadas pelo governo (Associação Nacional do Transporte de Cargas e Logística, Instituto Brasileiro de Geografia e Estatística IBGE, Comércio exterior COMEX), e dados concedidos publicamente por empresa privada (Agrolink), abrangendo variáveis cruciais como preços do milho e soja, valores de combustível, áreas colhidas e volumes exportados. A riqueza desses dados oferece uma oportunidade única para explorar as complexas interações que moldam o comportamento do mercado agrícola.

3.3 MATERIAIS E MÉTODOS

3.3.1 Base de dados

Para fundamentar a análise desta pesquisa, foram obtidos dados da Associação Nacional do Transporte de Cargas e Logística (ANTP), uma fonte reconhecida pela sua abrangência e confiabilidade no contexto logístico. A coleta inicial compreendeu o período de julho de 2001 a fevereiro de 2023, resultando em um total de 37.871 registros.

Para a análise detalhada dos fatores econômicos que influenciam o mercado agrícola, dados essenciais sobre o valor da saca de milho e soja foram diretamente coletados do Agrolink. Os dados filtrados diretamente pelo site foram para abranger apenas as informações relevantes para os estados do Paraná e Mato Grosso, refletindo a ênfase desta pesquisa.

O conjunto de dados do Agrolink, inicia-se em março de 2004 e indo até fevereiro de 2023, totalizando 944 registros.

Além das fontes previamente mencionadas, os dados de produção de milho e soja foram obtidos através do Sistema IBGE de Recuperação Automática (SIDRA). Essa base, vinculada IBGE, é uma referência confiável para informações estatísticas.

A filtragem dos dados na base SIDRA foi utilizada para que apenas os estados do Paraná e Mato Grosso e produtos de interesse, milho e soja, fossem incorporadas

à pesquisa, resultando em um total de 948 registros. Esses registros abrangem o período de janeiro de 2010 a fevereiro de 2023.

Outra fonte de dados utilizada nesta pesquisa provém do Sistema de Comércio Exterior (Comex), especificamente focada em avaliar o total exportado. A coleta desses dados envolveu uma filtragem direta do site, utilizando a Nomenclatura Comum do Mercosul (NCM), garantindo uma precisão substancial nas informações. Abaixo o quadro 6, mostra os códigos e suas especificações

Quadro 4: NCM

Código NCM	Descrição
07104000	Milho doce congelado, não cozido ou cozido em água ou vapor, congelados.
07099911	Milho doce, fresco ou refrigerado, para sementeira.
7099919	Milho doce, fresco ou refrigerado, exceto para sementeira.
7104000	Milho doce congelado, não cozido ou cozido em água ou vapor, congelados.
10051000	Milho para sementeira.
10059010	Milho em grão, exceto para sementeira.
10059090	Milho, exceto em grãos.
12010010	Soja para sementeira.
12010090	Outros grãos de soja, mesmo triturados.
12011000	Soja, mesmo triturada, para sementeira
12019000	Soja, mesmo triturada, exceto para sementeira.

Fonte: Autor (2023)

Os dados do Comex relacionados ao total exportado de milho e soja abrangem o período de janeiro de 2010 a outubro de 2023, totalizando inicialmente 9.746 registros. Após um cuidadoso tratamento que envolveu a categorização e agrupamento das NCM, a base foi refinada para 332 registros.

A última fonte de dados a ser incorporada ao escopo desta pesquisa consistirá em uma série histórica do dólar, obtida da plataforma *Investing*, uma plataforma online dedicada a fornecer informações e serviços relacionados a investimentos e finanças. Ele oferece uma ampla variedade de recursos, incluindo cotações em tempo real de ações, commodities, moedas, índices e outros instrumentos financeiros. Integrar essa fonte de dados ao conjunto já diversificado reforça a abordagem holística desta pesquisa, visando oferecer uma análise mais completa e precisa do cenário agrícola, considerando os fatores cambiais como parte integrante desse complexo ecossistema.

3.3.2 Ferramentas

Para realizar uma análise estatística detalhada da base de dados compilada, empregamos uma abordagem combinada, utilizando tanto o Microsoft Excel quanto a linguagem de programação Python.

O Excel foi instrumental na fase inicial da análise, permitindo uma exploração visual dos dados. A interface amigável do Excel proporcionou uma rápida compreensão das características gerais dos conjuntos de dados.

A linguagem de programação Python foi empregada para filtrar grandes quantidades de dados bem como para realizar análises estatísticas mais avançadas e personalizadas. Implementamos análises descritivas mais complexas, identificamos padrões e conduzimos visualizações detalhadas.

3.4 PROCEDIMENTOS METODOLÓGICOS

A Figura 8 ilustra os passos metodológicos a serem executados para atingir os objetivos específicos delineados nesta pesquisa. Essa abordagem visa, por conseguinte, alcançar o objetivo geral estabelecido.

Figura 8: CARACTERIZAÇÃO DA PESQUISA



Fonte: Autor (2023)

4 PROCESSAMENTO E ANÁLISE DOS RESULTADOS

4.1 SELEÇÃO DOS DADOS

Na busca por uma análise abrangente e representativa, optou-se por selecionar 18 atributos específicos dentre as quatro bases de dados utilizadas. O quadro 5 traz mais detalhes. Os critérios adotados para a seleção dos atributos basearam-se na sua relevância para os objetivos específicos da pesquisa.

Quadro 5: ATRIBUTOS SELECIONADOS

BASE	ATRIBUTO	DESCRIÇÃO	EXEMPLO
ANTP	MÊS	Data referente ao mês da cotação de combustível	01/01/2001
ANTP	ESTADO	Estado realizado a cotação de combustível	PARANA
ANTP	PREÇO MÉDIO REVENDA	Preço Médio nos postos	R\$ 2,50
Agrolink	MES_ANO	Data referente ao mês da cotação dos produtos	01/01/2001
Agrolink	UF	Estado realizado a cotação dos produtos	PR
Agrolink	PROD	Nome do produto	MILHO

Agrolink	RS_SACA	Valor médio da cotação do produto	R\$ 60,00
Agrolink	MED_NAC	valor médio nacional do produto	R\$ 60,00
SIDRA	DATA	Mês de referência de produção	Janeiro 2010
SIDRA	UF	Estado da produção	Paraná
SIDRA	PRODUTO	Produto listado	SOJA
SIDRA	PRODUÇÃO TOTAL	Total da produção do estado em tonelada	6299322
SIDRA	PRODUÇÃO NACIONAL	Total da produção do estado em tonelada	33424815
COMEX	ANO	Ano das exportações	2010
COMEX	MÊS_REF	Mês das exportações	01
COMEX	UF_PROD	Estado que se originou o produto	PR
COMEX	PRODUTO	Nome do produto	SOJA
COMEX	Quilograma Líquido	Peso líquido exportado	126370905

Fonte: Autor (2023)

4.2 PRÉ-PROCESSAMENTO E LIMPEZA DOS DADOS

Na continuidade do processamento e análise dos resultados, uma etapa fundamental foi a unificação de todos os atributos selecionados em uma única base de dados consolidada. Durante esse procedimento, houve a integração de dados semelhantes e a exclusão de registros duplicados. Foi realizada a padronização de atributos que apresentavam repetições em diferentes bases, assegurando coesão e uniformidade nos dados. Abaixo no quadro 6 consta a padronização.

Quadro 6: ATRIBUTOS CORRIGIDOS

ATRIBUTO	EXEMPLO
DATA	01/01/2012
UF	PR
PROD	MILHO

RS_SACA	R\$ 95,15
RS_NAC	R\$ 85,15
PRECO_COMB	R\$ 4,15
TOTAL_PROD	21036197,00
TOTAL_NAC	145021020,00
TON_EXP	3045989,6
DATA	R\$ 3,15

Fonte: Autor (2023)

Além disso, foi realizada a correção das unidades de medida para garantir uma comparação coerente. Por exemplo, valores originalmente expressos em quilogramas foram convertidos para toneladas, proporcionando uma uniformidade essencial para uma análise precisa e abrangente.

Durante o estágio de pré-processamento e limpeza dos dados, foi identificada uma lacuna específica na base da ANTP referente ao mês de setembro de 2022, atribuída à escassez de informações devido à pandemia de COVID-19. Em resposta a essa limitação, optou-se pela remoção dos dados correspondentes a esse período.

No atributo DATA, optou-se por excluir todas as datas anteriores a 2012, visando concentrar a análise em um período mais recente e relevante. Quanto ao atributo UF, foram mantidos apenas os estados do Mato Grosso (MT) e Paraná (PR), descartando-se as demais unidades federativas.

Considerando a natureza do setor de transporte de carga e combustível e tratores agrícolas, onde os veículos pesados, em sua maioria, utilizam diesel como combustível, a base foi filtrada para incluir apenas registros relacionados a esse tipo de combustível. Esse filtro visa assegurar que os dados analisados estejam alinhados com a realidade operacional dos veículos de carga

Essas ações foram conduzidas com o intuito de otimizar a qualidade e coerência dos dados utilizados na pesquisa, promovendo uma análise mais focalizada e alinhada aos objetivos estabelecidos.

4.3 TRANSFORMAÇÃO DOS DADOS

Para o atributo UF foi feita uma conversão dos dados textuais para transformá-los em dados numéricos como mostrado no quadro ().

Quadro 7 : TRANSFORMAÇÃO DE DADOS

ESTADO	NÚMERO REFERENTE
MT	1
PR	2

Fonte: Autor (2023)

4.4 ESTATÍSTICA DESCRITIVA

Após realizar a escolha, agrupamento, padronização e correção dos dados pertencentes a cada atributos. A base de dados consolidada para esta pesquisa foi refinada para um total de 532 registros, abrangendo 10 atributos essenciais, podemos analisar por meio do Quadro 8 os tipos de variáveis e suas variações de acordo com cada atributo.

Quadro 8: DESCRIÇÃO ATRIBUTOS

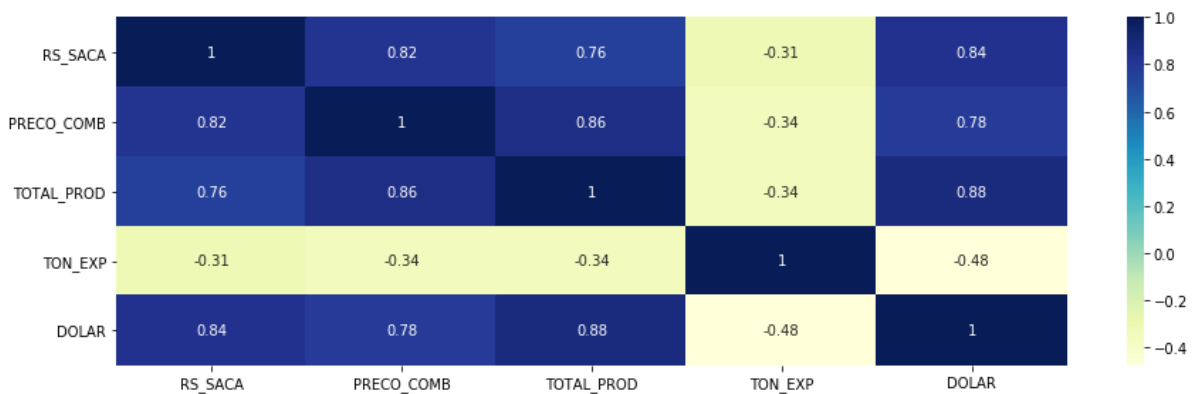
ATRIBUTO	VARIÁVEL	VARIAÇÃO
DATA	DATA	DE 2012 ATÉ FEVEREIRO DE 2023
UF	QUALITATIVA ORDINAIS	2 ESTADOS
PROD	QUALITATIVA NOMINAIS	2 PRODUTOS
RS_SACA	QUANTITATIVA CONTÍNUAS	DE R\$ 17,02 ATÉ R\$ 194,82
RS_NAC	QUANTITATIVA CONTÍNUAS	DE R\$ 19,60 ATÉ R\$ 192,00
PRECO_COMB	QUANTITATIVA CONTÍNUAS	DE R\$ 1,97 ATÉ R\$ 7,31
TOTAL_PROD	QUANTITATIVA CONTÍNUAS	DE 8641700 ATÉ 21036197
TOTAL_NAC	QUANTITATIVA CONTÍNUAS	DE 24311433,00 ATÉ 147545847
TON_EXP	QUANTITATIVA CONTÍNUAS	DE 0,18

		ATÉ 5361116,181
DOLAR	QUANTITATIVA CONTÍNUAS	DE R\$ 1, 71 ATÉ R\$ 5,66

Fonte: Autor (2023)

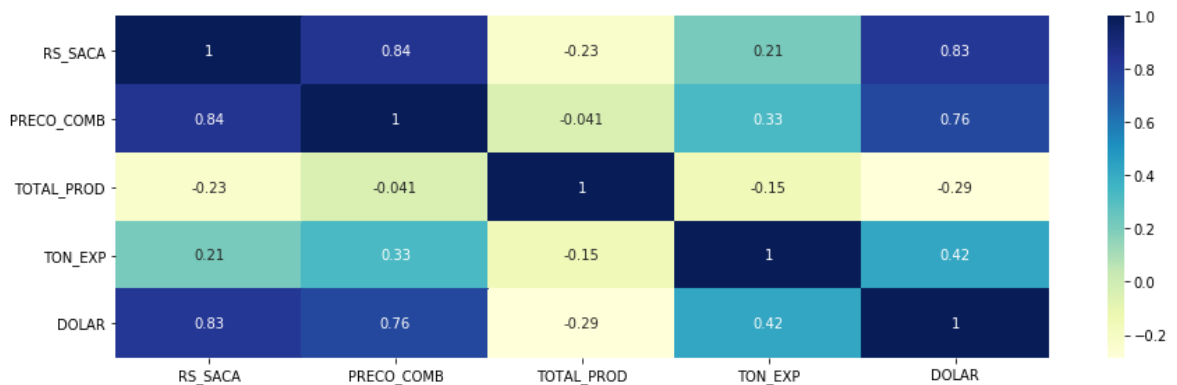
Na fase analítica foram exploradas as correlações entre atributos para compreensão do mercado agrícola entre os estados do Paraná e Mato Grosso. Para essa análise, foram selecionados cinco atributos cruciais, 'RS_SACA', 'PRECO_COMB', 'TOTAL_PROD', 'TON_EXP' e 'DOLAR'. Essa escolha estratégica visa examinar as relações temporais, os impactos dos preços das safras ('RS_SACA') e dos combustíveis ('PRECO_COMB'), a produção total ('TOTAL_PROD'), o volume exportado ('TON_EXP') e a influência da taxa de câmbio ('DOLAR'). Abaixo segue os gráficos referente matriz de correlações:

Gráfico 1: CORRELAÇÃO DOS ATRIBUTOS NO MATO GROSSO E NO MILHO



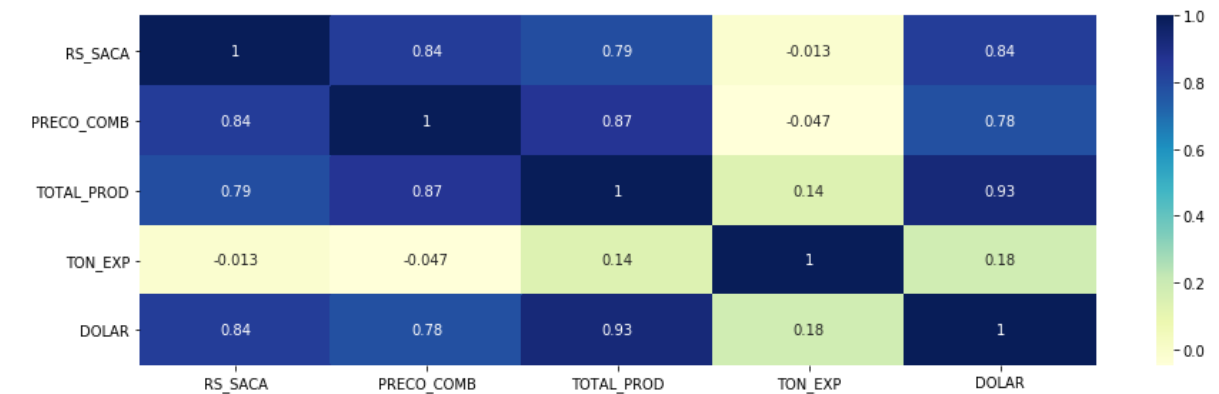
Fonte: Autor (2023)

Gráfico 2: CORRELAÇÃO DOS ATRIBUTOS NO PARANÁ E NO MILHO



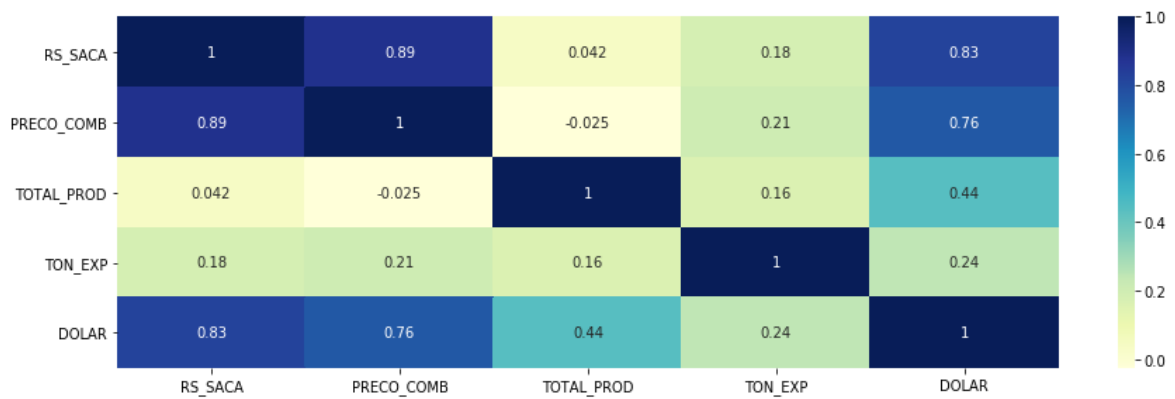
Fonte: Autor (2023)

Gráfico 3: CORRELAÇÃO DOS ATRIBUTOS NO MATO GROSSO E NA SOJA



Fonte: Autor (2023)

Gráfico 4: CORELAÇÃO DOS ATRIBUTOS NO PARANÁ E NA SOJA



Fonte: Autor (2023)

A análise de correlações entre os atributos nos estados do Paraná (PR) e Mato Grosso (MT) revelou padrões significativos nos relacionamentos entre as variáveis.

No Paraná, tanto para a produção de milho quanto para a de soja, as correlações mais expressivas foram observadas entre o preço da saca e o combustível, atingindo 0.84 e 0.89, respectivamente. Além disso, a relação entre o preço da saca e o dólar também foi marcante, registrando uma forte correlação de 0.83 em ambos os casos. A correlação entre o combustível e o dólar foi notável, atingindo 0.76.

No Mato Grosso, a produção de soja revelou correlações robustas, especialmente entre o preço da saca e o combustível (0.84) e entre o preço da saca e o dólar (0.84) que foi igual ao milho. Surpreendentemente, a produção de milho exibiu correlações mais marcantes, destacando-se a relação entre o preço da saca e o combustível (0.82). Além disso, a produção de milho apresentou correlações notáveis entre o combustível e a produção (0.86) e entre o dólar e a produção (0.93).

Uma observação relevante é que, em ambos os estados, a produção exportada ('TON_EXP') demonstrou menor correlação em comparação com outros atributos, sugerindo uma relação menos direta com os fatores analisados. As correlações robustas entre os preços das safras, os custos dos combustíveis e a taxa de câmbio ressaltam a sensibilidade desses setores às condições econômicas e cambiais. Estes resultados corrigidos proporcionam uma base sólida para análises mais aprofundadas, visando uma compreensão mais completa do comportamento do mercado agrícola entre o Paraná e Mato Grosso.

Prosseguindo com a análise, foi aprofundado a compreensão dos padrões identificados por meio da implementação de testes neurais. Essa abordagem avançada permite explorar relações mais complexas e não-lineares entre os atributos, proporcionando insights mais profundos sobre o comportamento do mercado agrícola nos estados. Ao utilizar redes neurais, visamos capturar nuances e padrões que podem não ser facilmente identificados por métodos convencionais.

Para continuar análise foram retirados das análises os atributos 'TOTAL_PROD' e 'TON_EXP'. Por terem correções mais baixas que as demais. Além do descarte das variáveis, foi realizada a reestruturação da base de dados de modo a poder ser feita a rede neural. Para isso os atributos 'RS_SACA', 'RS_NAC', foram divididos por produto e renomeadas para VALOR_M, VALOR_S, VALOR_NM e VALOR_NS.

A concepção da rede neural para análise envolveu a criação de um atributo. Este atributo, elaborado, tem como propósito calcular a probabilidade de os produtos serem exportados, introduzindo uma dimensão de previsibilidade nas decisões. Operando numa lógica de classificação binária, atribuímos o valor 1 para indicar a

destinação ao mercado interno e o valor 2 para indicar uma inclinação favorável à exportação. Ao integrar esse atributo à arquitetura da rede neural, buscamos não apenas prever se um produto será destinado ao mercado interno ou à exportação, mas também compreender as influências e determinantes por trás dessas decisões.

Optado por utilizar o algoritmo *perceptron* com o objetivo específico de analisar os pesos relativos de cada atributo. Esta abordagem ofereceu uma visão detalhada das interações entre os diferentes fatores que influenciam a probabilidade de um produto ser exportado dentro os estados.

Os resultados iniciais do teste geral da base não revelaram melhorias significativas, levando à decisão de uma abordagem mais detalhada. Diante disso, optamos por estudar individualmente o comportamento de cada produto em cada estado. Essa estratégia visa identificar padrões específicos e nuances associadas a cada produto, aprimorando a capacidade de previsão e interpretação da rede neural.

Para conduzir os estudos, utilizamos uma taxa de aprendizado de 70%, o que representa a proporção dos dados usados para treinar a rede neural. Além disso, reservamos 20% da base de dados para a fase de teste, garantindo uma avaliação robusta do desempenho do modelo em condições diversas.

Neste primeiro estágio, foram realizados os mesmos parâmetros de testes para ambas as análises dos estados. Abaixo consta quadro com os significados de cada peso estão os resultados obtidos respectivamente:

Quadro 9: SIGNIFICADO DOS PESOS

Peso	Significado
P1	Preço do produto no estado
P2	Média Nacional do valor do produto
P3	Preço do Combustível no Estado
P4	Cotação do Dólar

Fonte: Autor (2023)

Mato Grosso - MILHO:

Taxa de acerto: 0,42

Pesos Finais:

p1 = 819.452

p2 = 1094.883

p3 = 123.8784
p4 = 116.390075

Mato Grosso - SOJA:

Taxa de acerto: 0,42
Pesos Finais:
p1 = 2727.897
p2 = 2686.298
p3 = 129.8802
p4 = 120.675195

Paraná - MILHO:

Taxa de acerto: 0,88
Pesos Finais:
p1 = 620.022
p2 = 604.267
p3 = 38.1571
p4 = 42.940965

Paraná - SOJA:

Taxa de acerto: 0,8461538461538461
Pesos Finais:
p1 = 1069.681
p2 = 1097.004
p3 = 35.9878
p4 = 40.2523

Geralmente, pesos mais altos indicam uma maior influência da variável na decisão do modelo. Em uma perspectiva geral, parece que os valores dos produtos ('VALOR_M' e 'VALOR_NM') têm uma influência substancial nas decisões do modelo, seguidos pelos preços do combustível ('PRECO_COMB'). O dólar (DOLAR) tem uma

influência menor, mas ainda contribui para o processo decisório. Estas observações podem orientar ajustes no modelo para otimizar seu desempenho.

Prosseguiu a análise utilizando o pacote Keras, uma biblioteca de aprendizado profundo amplamente utilizada, devido à sua eficácia e flexibilidade na construção e treinamento de redes neurais. Optado por reservar 30% dos dados para teste, visando avaliar o desempenho do modelo em uma amostra independente, garantindo uma avaliação robusta de sua capacidade de generalização para dados não vistos.

Ao definir a arquitetura do modelo no Keras, escolhemos a função de ativação 'linear'. Essa escolha foi feita considerando o tipo de problema em questão, onde estamos interessados em uma previsão contínua em vez de uma classificação discreta. A ativação 'linear' é apropriada para problemas de regressão, onde a saída deve ser uma estimativa numérica.

No processo de compilação do modelo, foi optado por usar a função de perda (*loss*) 'mse' (*Mean Squared Error*) e o otimizador 'adam'. A escolha da função de perda 'mse' é adequada para problemas de regressão, pois mede o erro médio quadrático entre as previsões do modelo e os valores reais. Quanto ao otimizador 'adam', ele é uma escolha comum para problemas de regressão devido à sua eficácia e eficiência em adaptar automaticamente a taxa de aprendizado durante o treinamento, o que pode acelerar a convergência do modelo.

A escolha da métrica 'accuracy' no contexto da análise binária é fundamentada na natureza do problema que estamos abordando. Ao lidar com atributos binários, como a decisão de exportar ou não um produto, a métrica 'accuracy' torna-se relevante para avaliar a precisão global do modelo em classificar corretamente esses resultados binários.

Essas escolhas foram orientadas pela natureza do problema de previsão contínua que estamos enfrentando e pelas características específicas dos dados em questão, visando uma modelagem precisa e eficaz.

Para a construção do modelo, optamos por utilizar uma arquitetura composta por 5 neurônios, cada uma contribuindo para diferentes aspectos da análise, com uma única saída. A escolha de empregar múltiplas redes neurais visa capturar de maneira mais abrangente as nuances presentes nos dados, permitindo que o modelo se adapte a diferentes padrões e complexidades subjacentes.

Ao implementar a arquitetura do modelo, definimos alguns parâmetros essenciais para o treinamento da rede neural. Foi estabelecido o número de épocas (*epochs*) como 200, indicando o número de iterações completas sobre o conjunto de dados durante o treinamento. Além disso, foi especificado o tamanho do lote (*batch_size*) como 32. Este parâmetro determina a quantidade de exemplos de treinamento utilizados em cada iteração do algoritmo de otimização. A escolha de um tamanho de lote moderado, como 32, equilibra eficiência computacional e estabilidade do treinamento.

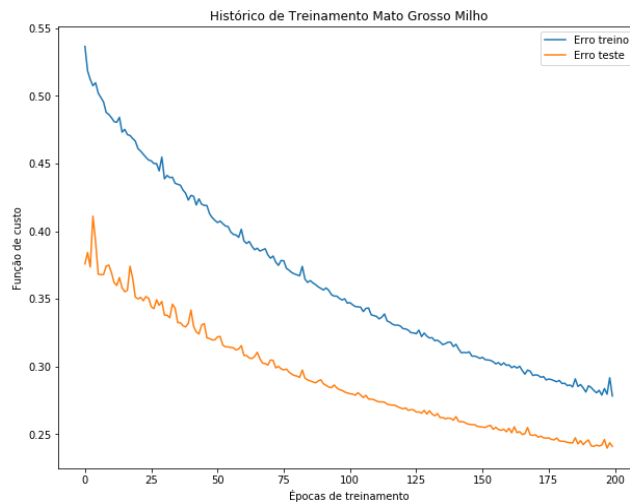
Ao avaliar o desempenho do modelo, duas métricas emergem como indicadores cruciais: *Loss* (Perda) e *Accuracy* (Acurácia). Essas métricas são essenciais para compreender como o modelo está se saindo em termos de precisão e capacidade de aprendizado. *Loss* (Perda): *Loss* representa a discrepância entre as previsões do modelo e os valores reais. Para problemas de regressão, como é o caso do valor de Milho e Soja, utilizamos a *Mean Squared Error* (Erro Quadrático Médio) como medida de perda. *Accuracy* (Acurácia): A acurácia é uma métrica crucial em problemas de classificação binária, como a decisão de exportar ou não um produto.

Ao examinar os resultados do modelo de redes neurais aplicado aos dados do mercado agrícola nos estados, uma análise mais aprofundada revela nuances significativas em relação à capacidade preditiva e generalização do modelo.

No cenário mato-grossense, as complexidades variam entre os produtos estudados. No caso do milho, o modelo apresenta uma razoável concordância entre previsões e valores reais, evidenciada por uma classe de 0.4494. Contudo, a acurácia de 0.4318 sugere que o modelo enfrenta desafios em prever com precisão o valor desse produto específico. Ao considerar a soja, os desafios são mais pronunciados, refletidos em uma perda substancial de 3.2997 e uma acurácia de 0.4545, indicando

dificuldades significativas em capturar as nuances desse mercado. Na fase de validação, o modelo demonstra uma capacidade notável de generalização para novos dados, conforme atestam as métricas de loss (0.5147) e acurácia (0.4474). Esses resultados sugerem uma adaptabilidade consistente do modelo diante de diferentes conjuntos de dados, o que é crucial para sua eficácia prática. Abaixo segue os gráficos 5 e 6 referentes ao desempenho do modelo.

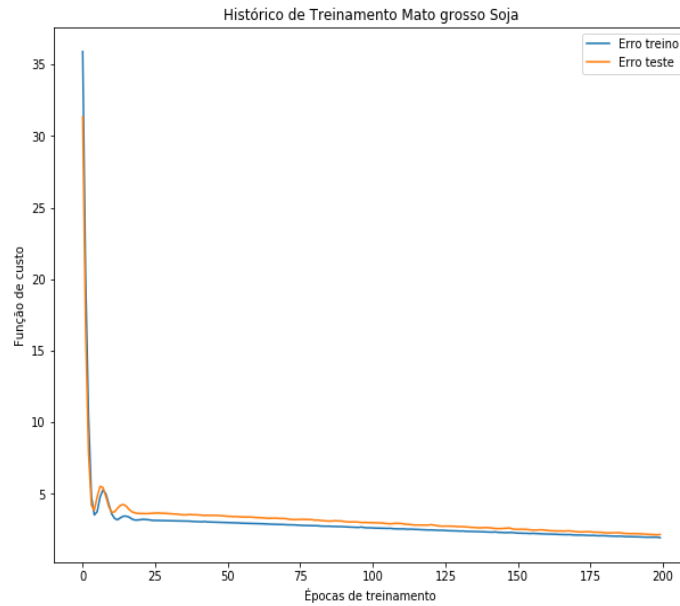
Gráfico 5: HISTORICO DE TREINAMENTO MATO GROSSO MILHO



Fonte: Autor (2023)

A análise do gráfico revela uma diferença notável entre os conjuntos de treinamento e teste. Entretanto, é crucial observar que a análise apresenta uma certa instabilidade, refletida na natureza irregular do gráfico, caracterizada por variações abruptas e descidas. Essas oscilações podem indicar uma sensibilidade do modelo a certas condições específicas ou a presença de outliers que afetam sua performance em certos contextos.

Gráfico 6: HISTORICO DE TREINAMENTO MATO GROSSO SOJA



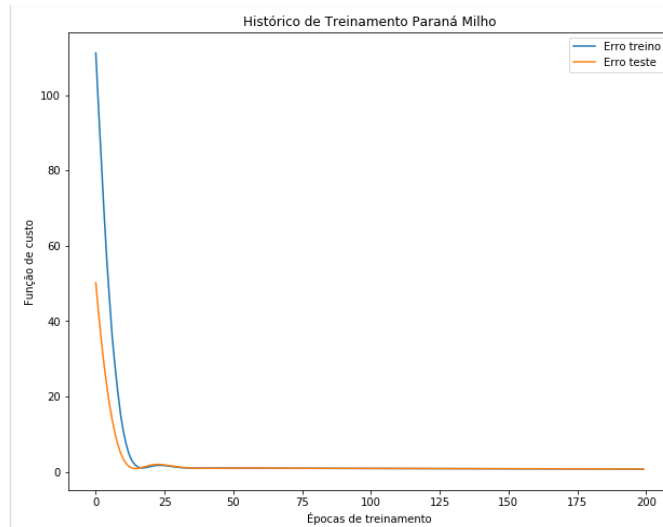
Fonte: Autor (2023)

A análise da soja revela uma dinâmica interessante entre os conjuntos de treinamento e teste. Essa disparidade entre treinamento e teste pode indicar um possível ajuste do modelo aos dados de treinamento específicos, dificultando sua capacidade de lidar eficientemente com novos dados. O conjunto de treinamento apresentou um desempenho superior em comparação com o conjunto de teste, uma observação que merece atenção. Durante o treinamento, observou-se uma redução significativa na função de custo já na primeira época, indicando uma rápida adaptação do modelo aos padrões presentes nos dados de treinamento.

No Paraná, os resultados do modelo são mais promissores, especialmente para o milho. Aqui, observamos uma loss de 1.4500 e uma acurácia de 0.8095 durante o treinamento, indicando uma concordância robusta entre previsões e valores reais. Essa eficácia é corroborada pela validação, que apresenta uma perda de 1.3941 e uma acurácia de 0.7812, ressaltando a habilidade do modelo em generalizar para novos dados. Quanto à soja no Paraná, o modelo revela um desempenho ainda mais sólido, com uma perda de 0.6995 e uma acurácia de 0.7302 durante o treinamento. Na fase de validação, essas métricas mantêm sua robustez, registrando uma perda de 0.7283 e uma acurácia de 0.8438. Esses números indicam uma capacidade consistente do modelo em generalizar para diferentes conjuntos de dados, apontando

para sua efetividade prática. Abaixo seguem os gráficos 7 e 8 referentes ao desempenho do modelo.

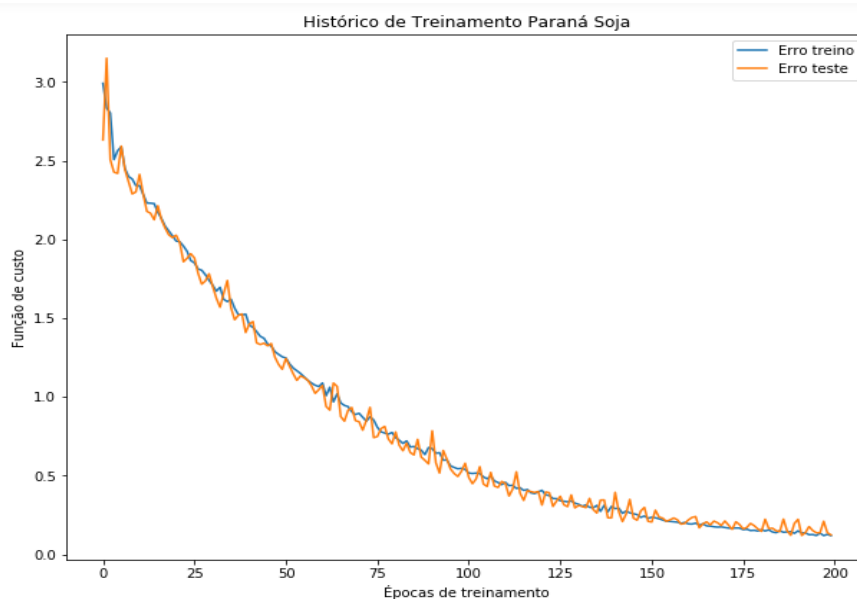
Gráfico 7: HISTORICO DE TREINAMENTO PARANÁ MILHO



Fonte: Autor (2023)

O alinhamento subsequente entre os gráficos de treinamento e teste, mantendo-se estáveis ao longo das épocas seguintes, sugere que o modelo conseguiu generalizar efetivamente para novos dados. Essa consistência entre os conjuntos indica um bom desempenho do modelo na previsão do cultivo de milho no Paraná, com a capacidade de aplicar padrões aprendidos durante o treinamento para dados não vistos.

Gráfico 8: HISTORICO DE TREINAMENTO PARANÁ SOJA



Fonte: Autor (2023)

A análise revela uma dinâmica distinta em comparação com os cenários anteriores. Observa-se que a redução na função de custo durante o treinamento foi mais gradual e linear. Em contraste, no conjunto de teste, a curva apresenta oscilações, subindo e descendo em torno de valores específicos. Essa discrepância entre o treinamento e o teste sugere uma possível dificuldade do modelo em generalizar adequadamente para os dados de teste

5. CONCLUSÃO

A interseção entre Gestão da Informação, mineração de dados e agricultura representa um avanço no entendimento e otimização dos processos agrícolas. Ao capacitar profissionais na coleta, organização e análise eficiente de dados, o curso proporciona as habilidades necessárias para explorar insights valiosos que orientam decisões estratégicas no setor agrícola. A agricultura moderna é intrinsecamente dependente de dados, desde as tendências de mercado até as variáveis climáticas e econômicas. Portanto, a expertise em Gestão da Informação não apenas aprimora a eficiência operacional, mas também promove a sustentabilidade e a inovação na agricultura, garantindo a utilização inteligente de recursos e impulsionando a produtividade em consonância com as demandas dinâmicas do setor.

A presente pesquisa proporcionou uma análise abrangente do mercado agrícola nos estados do Mato Grosso e Paraná, revelando nuances e particularidades que delineiam a dinâmica dessas regiões. Ao longo deste trabalho, cumprimos integralmente nosso objetivo geral, que consistia em analisar o comportamento desse mercado por meio da mineração de dados. Os resultados obtidos permitiram uma compreensão mais profunda dos fatores que moldam esses cenários agrícolas específicos.

Os resultados da análise evidenciam as distinções no comportamento dos estados do Mato Grosso e Paraná no mercado agrícola. A correlação aponta que, no Paraná, há uma ligação mais robusta entre três fatores principais: o preço da saca, o preço do combustível e a cotação do dólar. Por outro lado, no Mato Grosso, os fatores apresentam correlações mais diversificadas e sinérgicas, indicando uma dinâmica única. Surpreendentemente, ao empregar a rede neural, o estudo revelou resultados mais significativos no Paraná do que no Mato Grosso, destacando a complexidade e nuances distintas desses mercados agrícolas regionais.

A análise crítica das métricas de *loss* e acurácia oferece uma visão mais refinada da eficácia do modelo, destacando áreas de sucesso e desafios específicos. Esses resultados não apenas contribuem para a compreensão do comportamento do mercado agrícola em nível regional, mas também fornecem insights valiosos para refinamentos e melhorias futuras no modelo. Essa abordagem crítica e detalhada enriquece a compreensão do papel das redes neurais na análise do mercado agrícola, especialmente nos estados do Mato Grosso e Paraná.

No entanto, é crucial reconhecer as limitações deste estudo. A ausência de dados de exportação, como discutido anteriormente, destaca a necessidade de uma abordagem mais abrangente na coleta de informações, considerando possíveis variações nas classificações de produtos e rotas de exportação. Essas lacunas podem existir devido a diferentes classificações de produtos (NCMs) ou à entrada por meio de outros estados, ressaltando a importância de uma abordagem mais holística na obtenção de dados futuros.

Dessa forma, recomendamos que estudos futuros aprofundem essas lacunas, incorporando dados adicionais e considerando fatores externos não explorados neste trabalho. A inclusão de variáveis como condições climáticas, políticas governamentais e eventos econômicos pode enriquecer ainda mais a compreensão do comportamento do mercado agrícola.

Em síntese, este trabalho contribuiu para a análise do mercado agrícola nos estados do Mato Grosso e Paraná, demonstrando as potencialidades da mineração de dados nesse contexto. Ao enfrentar desafios e reconhecer áreas para aprimoramento, abre-se caminho para pesquisas futuras mais abrangentes e aprofundadas nesse campo crucial para a economia brasileira.

6 REFERÊNCIAS

Agência Nacional do Petróleo e Gás Natural e Biodiesel. Série Histórica de Preços de Combustíveis e de GLP. Disponível em: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/serie-historica-de-precos-de-combustiveis>. Acesso em: 20 nov. 2023.

AGROLINK. Cotações. Disponível em: <https://www.agrolink.com.br/cotacoes/>. Acesso em: 20 nov. 2023.

BUAINAIN, Antônio Márcio; FILHO, Hildo Meirelles de Souza. A política agrícola no Brasil: evolução e principais instrumentos. In : BATALHA, Mário Otávio . Gestão Agroindustrial : GEPAL: **Grupo de estudo e pesquisas agroindustriais**. 4. ed. – 2.reimpr., São Paulo: Atlas, 2007.

BEAL, A. **Gestão Estratégica da Informação**: como transformar a informação e a tecnologia da informação em fatores de crescimento e de alto desempenho nas organizações. 3. ed. São Paulo: Atlas, 2008.

BOENTE, Alfredo Nazareno Pereira; OLIVEIRA, Fabiano Saldanha Gomes de; ROSA, José Luiz dos Anjos. **Utilização de ferramentas de kdd para integração de aprendizagem e tecnologia em busca da gestão estratégica do conhecimento na empresa**. 12 f. Monografia (Especialização) - Centro Universitário Estadual da Zona Oeste, Rio de Janeiro. Disponível em: https://www.aedb.br/seget/arquivos/artigos07/1219_Artigo%20SEGET%202007.pdf. Acesso em: 28 jun 2023

CASTRO, L. N. de; FERRARI, D. G. **Introdução à mineração de dados**: conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

DAVENPORT, T. H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da Informação. São Paulo: Futura, 1998.

DOURADO, Vania Maria Celestino. **Agronegócio brasileiro**: um potencial econômico. 2005. 58 f. Monografia (Graduação) – Faculdade de Ciências Sociais Aplicadas, Centro Universitário de Brasília, Brasília, 2005. Disponível em: <https://repositorio.uniceub.br/jspui/handle/235/10985>. Acesso em: 25 jun 2023

DUARTE, Jason de Oliveira; MATTOSO, Marcos Joaquim; GARCIA, João Carlos. Importância Socioeconômica: milho. Milho. 2021. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/milho/pre-producao/socioeconomia/importancia-socioeconomica>. Acesso em: 30 jun. 2023.

EMBRAPA. VII Plano Diretor da Embrapa 2020-2030. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1126091/vii-plano-diretor-da-embrapa-2020-2030>. Acesso em: 1 jun. 2023.

EMBRAPA (comp.). Soja em números (safra 2021/22). Disponível em: <https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>. Acesso em: 30 jun. 2023.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From **Data Mining to Knowledge Discovery in Databases**. AI Magazine, [S. l.], v. 17, n. 3, p. 37, 1996. DOI: 10.1609/aimag.v17i3.1230. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1230>. Acesso em: 25 jun. 2023.

Ferra Z, Ana Paula. **MINERAÇÃO DE PROCESSOS EM UMA BASE DE TRANSAÇÕES DE ATENDIMENTO AO CLIENTE EM UM COMÉRCIO ELETRÔNICO**. 2022. 53 f. TCC (Graduação) - Curso de Gestão da Informação, Universidade Federal do Paraná, Curitiba, 2022.

FIGUEIREDO, Margarida Garcia de; BARROS, Alexandre Lahós Mendonça de; GUILHOTO, Joaquim José Martins. Relação econômica dos setores agrícolas do Estado do Mato Grosso com os demais setores pertencentes tanto ao Estado quanto ao restante do Brasil. **Revista de Economia e Sociologia Rural**, [S.L.], v. 43, n. 3, p. 557-575, set. 2005. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-20032005000300008>.

GOLDSHMIDT, R. O.; PASSOS, E. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações**. Rio de Janeiro: Elsevier, 2005. ISSN 07102755.

INVESTING. USD/BRL - Dólar Americano Real Brasileiro. Disponível em: <https://br.investing.com/currencies/usd-brl-historical-data>. Acesso em: 20 nov. 2023.

MARTINS, Talita Mauad. **Ciclos e previsão cíclica dos preços das commodities: um modelo de indicador antecedente para a commodity açúcar**. 2009. 151 p. São Paulo. Disponível em: <https://www.teses.usp.br/teses/disponiveis/12/12139/tde-05022010-122747/pt-br.php> Acesso em: 29 mai. 2023

Ministerio da industria comercio exterior e serviços. Exportação e Importação Geral. Disponível em: <http://comexstat.mdic.gov.br/pt/home>. Acesso em: 20 nov. 2023.

Ministério do Desenvolvimento Indústria Comércio e Serviços. **Balança Comercial e Estatísticas de Comércio Exterior**. Disponível em: <https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior/estatisticas>. Acesso em: 01 jun. 2023.

MOURA, J.G.; CÂMARA, S.F.; LIMA, R.C. Eficiência alocativa e crescimento econômico: o papel do setor agrícola (compact disc). In: CONGRESSO **BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL**, 41., Foz do Iguaçu, 1999. Anais. Brasília: SOBER, 1999.

NANTES, José Flavio Diniz; SCAERPELLI, Moacir. Elementos de gestão na produção rural; FILHO, Hildo Meirelles de Souza. Desenvolvimento agrícola sustentável. In: BATALHA, Mário Otávio. Gestão Agroindustrial : GEPAL: **Grupo de estudo e pesquisas agroindustriais**. 3. Ed. – 2. Reimpr. , São Paulo: Atlas, 2008

NOGUEIRA, E. D. A. **Avaliação de brand equity sob a perspectiva do consumidor nas mídias sociais por meio da mineração de opinião e análise de redes sociais**. 2015. 237 f. DISSERTAÇÃO (Mestrado) – Programa de Pós-graduação em Ciência, Gestão e Tecnologia da Informação, Ciências Sociais Aplicadas, Universidade Federal do Paraná. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/41257/R%20-%20D%20-%20EDUARDO%20DIMAS%20ANDRINO%20NOGUEIRA.pdf?sequence=2&isAllowed=y> . Acesso em: 28 jun. 2023

ROESSING, A. C.; SANCHES, A. C.; MICHELLON, E.; As Perspectivas de Expansão da Soja. Anais dos Congressos. XLIII Congresso da Sober em Ribeirão Preto. São Paulo, 2005. Disponível em:

<https://www.yumpu.com/pt/document/view/31098284/anais-do-xliv-congresso-da-sociedade-sober>. Acesso em 30 jun. 2023.

RONCON, Natalia. **A IMPORTÂNCIA DO SETOR AGRÍCOLA PARA A ECONOMIA BRASILEIRA**. 2011. 71 f. TCC (Graduação) - Curso de Administração de Empresas, Fundação Educacional do Município de Assis, Assis, 2011. Disponível em: <https://cepein.femanet.com.br/BDigital/arqTccs/0811260631.pdf>. Acesso em: 19 jun. 2023.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio**. 4 ed. Porto Alegre: Bookman, 2019.

SIDRA. Levantamento Sistemático da Produção Agrícola. Disponível em: <https://sidra.ibge.gov.br/tabela/6588>. Acesso em: 20 nov. 2023.

SILVA, Ariana Cericatto da; LIMA, Érica Priscilla Carvalho de; BATISTA, Henrique Rogê. A IMPORTÂNCIA DA SOJA PARA O AGRONEGÓCIO BRASILEIRO: UMA ANÁLISE SOB O ENFOQUE DA PRODUÇÃO, EMPREGO E EXPORTAÇÃO. **Encontro de Economia Catarinense**, Florianópolis, v. 5, n. 5, p. 1-21, 1 jul. 2011. Disponível em: <https://apec.pro.br/anais/v-eeec/anais/4-EEC%202011.PDF>. Acesso em: 29 jun. 2023.

SILVA, Leandro Augusto; PERES, Sarajane M.; BOSCARIOLI, Clodis. **Introdução à Mineração de Dados - Com Aplicações em R**. [Digite o Local da Editora]: Grupo GEN, 2016. 9788595155473. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788595155473/>. Acesso em: 25 jun. 2023.

SOUZA, Angela Ferreira de. **ANÁLISE DE DADOS EM BASE DA GUARDA MUNICIPAL DE CURITIBA**. 2021. 159 f. TCC (Graduação) - Curso de Gestão da Informação, Universidade Federal do Paraná, Curitiba, 2021

Souza, D. K. F., Silveira, R. L. F., & Ballini, R. (2023). Efeito da expansão da safra de inverno de milho no Brasil sobre a sazonalidade dos preços spot. *Revista de Economia e Sociologia Rural*, 61(4), e262824. <https://doi.org/10.1590/1806-9479.2022.262824>

SYNGENTA. **Milho: maior cultura agrícola produzida no mundo**. Disponível em: <https://portal.syngenta.com.br/noticias/milho-maior-cultura-agricola-produzida-no-mundo/#:~:text=O%20milho%20%C3%A9%20a%20cultura,milh%C3%B5es%20de%20toneladas%20do%20gr%C3%A3o..> Acesso em: 30 jun. 2023.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data mining: Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009. 900 p.

5 APÊNDICE

5.1 APÊNDICE A - ATRIBUTOS DESCARTADOS

BASE	ATRIBUTO	MOTIVO
ANTP	REGIÃO	O atributo estado é mais específico
ANTP	NÚMERO DE POSTOS PESQUISADOS	A quantidade de postos não altera o resultado
ANTP	UNIDADE DE MEDIDA	Igual para todos os valores
ANTP	DESVIO PADRÃO REVENDA	Cálculo não utilizado na análise
ANTP	PRECO MÍNIMO REVENDA	Optou-se pela média
ANTP	PRECO MÁXIMO REVENDA	Optou-se pela média
ANTP	MARGEM MÉDIA REVENDA	O percentual de lucro dos postos não faz parte da análise
ANTP	COEF DE VARIAÇÃO REVENDA	Cálculo não utilizado na análise
ANTP	PRECO MÉDIO DISTRIBUIÇÃO	Valores que não interferem na análise
ANTP	DESVIO PADRÃO DISTRIBUIÇÃO	Valores que não interferem na análise
ANTP	PRECO MÍNIMO DISTRIBUIÇÃO	Valores que não interferem na análise
ANTP	PRECO MÁXIMO DISTRIBUIÇÃO	Valores que não interferem na análise
ANTP	COEF DE VARIAÇÃO DISTRIBUIÇÃO	Valores que não interferem na análise
COMEX	PAIS	Não mostrou relevância para análise
COMEX	Código NCM	Utilizado produto no lugar
COMEX	Descrição NCM	Utilizado produto no lugar
COMEX	Valor FOB (US\$)	Valores que não interferem na análise
SIDRA	SAFRA	Independente da safra o valor foi considerado.
SIDRA	COD_PRO	Utilizado produto no lugar
SIDRA	MEDIDA	Igual para todos os valores