

Análise da Comercialização de Grãos Usando a Metodologia CRISP-DM

Jaisson Duarte¹ and Jeam Carlos Braun²

¹Carlos Amaral Hobig

*210662@upf.br; 212196@upf.br; ...

Received: . Revised: . Accepted: .

1 Introdução

A comercialização de grãos é um processo complexo e de alto risco para produtores rurais, pois a volatilidade dos preços no mercado de commodities pode impactar diretamente o retorno financeiro. Fatores como a demanda global, condições climáticas e políticas econômicas expõem os produtores a um risco significativo de perda de receita. Nesse cenário, a análise de dados surge como uma ferramenta essencial para a tomada de decisões estratégicas, ajudando produtores e cooperativas a maximizar o lucro e garantir a sustentabilidade das operações.

O principal objetivo deste projeto, que aplica o modelo CRISP-DM, é utilizar a análise de dados históricos para identificar e compreender as oscilações do mercado de grãos. Ao analisar os padrões de cotação e comercialização, a pesquisa busca determinar o momento mais oportuno para a venda, minimizando riscos e otimizando os resultados financeiros.

2 Metodologia

Este projeto foi conduzido seguindo a metodologia CRISP-DM (Cross-Industry Standard Process for Data Mining). A aplicação do CRISP-DM permitiu uma abordagem sistemática, dividida em seis fases principais: Entendimento do Negócio, para definir o problema e os objetivos; Entendimento dos Dados, para coletar e explorar as informações; Preparação dos Dados, para limpar e organizar o dataset; Modelagem, para aplicar técnicas estatísticas e de machine learning; Avaliação, para validar os resultados obtidos; e Disseminação, para apresentar as conclusões e as recomendações.

2.1 Entendimento do Negócio

A comercialização de grãos é uma etapa crítica para produtores rurais, servindo como a principal via para obter

retorno financeiro. A volatilidade dos preços, influenciada por fatores globais e locais, expõe os produtores a um risco financeiro significativo. Uma decisão de venda inadequada, seja no momento ou no preço, pode levar à perda de receita. Diante desse desafio.

2.2 Entendimento dos Dados

A análise foi realizada com dados de comercialização de grãos da cooperativa, extraídos do sistema ERP, cobrindo o período de 2020 a 2024. O dataset consolidado inclui informações detalhadas sobre os eventos de depósito, como o Ano, a Cidade, o Tipo do Grão, a Cotação no Depósito e a Quantidade em Sacas. Dados de faturamento e estoque, incluindo a Cotação Atual, também foram integrados para uma visão completa.

Durante a avaliação da qualidade dos dados, foi identificado um desafio na variável "Cotação" devido a registros ausentes. Para mitigar o problema e preservar a integridade do dataset, os valores faltantes foram preenchidos com a média da cotação da cidade correspondente.

A análise exploratória revelou a alta volatilidade do mercado. Um exemplo notável foi a valorização da soja em Ibirubá, com a cotação média subindo 44,3% entre 2020 e 2024. Em contraste, o milho na mesma cidade apresentou uma desvalorização de 38% no mesmo período, reforçando a importância do monitoramento de mercado para a tomada de decisão.

2.3 Preparação dos Dados

Para esta análise, foram selecionadas as informações de cotação na época do depósito e a cotação atual. Essa escolha se deu por sua relevância direta para determinar a rentabilidade de cada operação. Para um estudo mais focado, o dataset original foi filtrado para incluir apenas as observações do grão trigo, com depósitos realizados no ano de 2023. Essa segmentação permitiu um estudo mais

aprofundado e relevante para o produtor.

2.4 Modelagem

Com o objetivo de prever o valor da cotação do trigo, a técnica de Regressão Linear foi selecionada. O modelo estabelece uma relação direta entre a variável de entrada, a Cotação no Depósito (variável independente), e a variável de saída, a Cotação Atual (variável dependente), que o modelo busca estimar.

2.5 Avaliação

A qualidade do modelo de regressão foi avaliada pela métrica R^2 , que indica a proporção da variância na cotação atual explicada pelo modelo. Para validar a relevância dos resultados no contexto do negócio, uma consulta foi realizada com os analistas da cooperativa. A avaliação validou a importância da análise da dinâmica de preços, reforçando a necessidade de os produtores monitorarem as oscilações do mercado para evitar perdas e garantir a rentabilidade.

2.6 Implantação

Apesar de o projeto ter gerado insights valiosos, a implantação das recomendações não foi realizada na cooperativa. A disseminação dos resultados se deu por meio da apresentação das análises e conclusões à equipe de negócio, visando validar a relevância do estudo e preparar o terreno para uma futura implementação. O projeto serviu como um estudo-piloto, comprovando o potencial da análise de dados para otimizar a comercialização de grãos.

3 Resultados

A análise dos dados da cooperativa foi organizada em três visões no Power BI, cada um focado em um aspecto da comercialização de grãos: Análise Geral do Mercado, Análise de Cotações e Faturamento, e a Análise de Regressão.

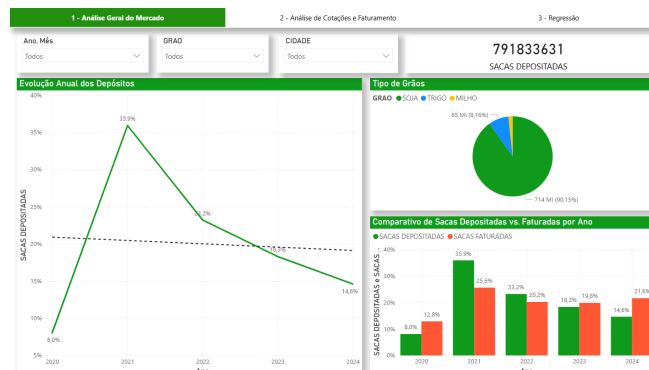


Figure 1: Análise Geral do Mercado

A primeira visão foca no volume de depósitos. O gráfico de linha "Evolução Anual dos Depósitos" mostra uma forte alta no volume em 2021, seguida de uma queda contínua até 2024. Isso pode indicar uma redução na safra ou uma mudança no comportamento dos produtores em relação à cooperativa. Já o gráfico de barras "Comparativo de Sacas Depositadas vs. Faturadas por Ano" revela a diferença entre o que foi depositado e o que foi de fato faturado em cada ano.

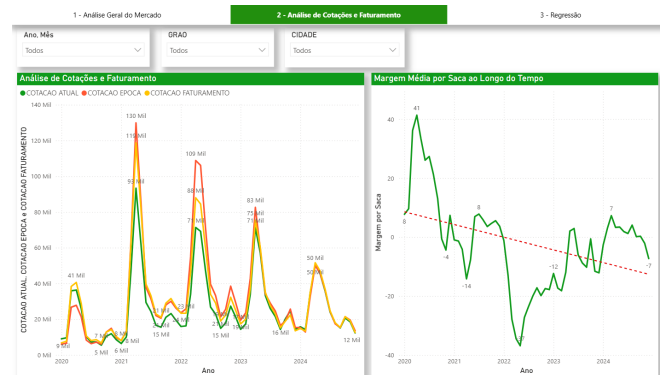


Figure 2: Análise de Cotações e Faturamento

A segunda visão apresenta o gráfico "Análise de Cotações e Faturamento" mostra as flutuações das cotações ao longo do tempo. É possível notar que, em alguns períodos, a Cotação de Faturamento (valor de venda) superou significativamente a Cotação de Depósito (valor de compra), indicando momentos de rentabilidade. O gráfico "Margem Média por Saca ao Longo do Tempo" quantifica essa variação, revelando períodos de margem positiva e negativa. A tendência geral (linha tracejada) mostra uma leve queda na margem ao longo dos anos, ressaltando a importância do planejamento de venda para mitigar perdas.

A terceira visão foi criada para diagnosticar o desempenho do modelo de regressão linear. Os cartões no topo do painel mostram que o R^2 e o R^2 ajustado são próximos de zero, confirmando que o modelo possui um baixo poder preditivo. Os gráficos de dispersão "Comportamento dos Resíduos" e "Resíduos vs. Valores

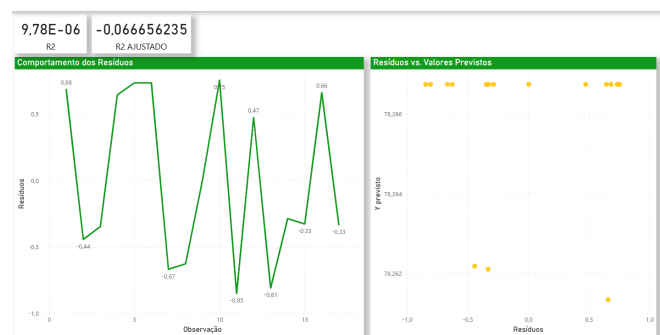


Figure 3: Análise de Regressão

Previstos" reforçam essa conclusão, mostrando uma distribuição aleatória e sem padrões dos erros. Esses resultados, em conjunto, indicam que a relação entre a variável de entrada e a de saída é fraca, e que o modelo, da forma como foi construído, não é ideal para prever a cotação futura.

References

4 Conclusão

Este projeto demonstrou que, apesar da alta volatilidade no mercado de grãos, a análise de dados é uma ferramenta poderosa para auxiliar na tomada de decisão. As análises realizadas no Power BI confirmaram a importância de monitorar de perto as flutuações de preços, destacando a necessidade de um planejamento de venda estratégico para otimizar a rentabilidade. O modelo de regressão linear, embora com baixo poder preditivo, validou a relação entre as cotações, mas evidenciou a necessidade de incluir mais variáveis para construir um modelo mais robusto. O projeto completo pode ser encontrado no github¹.

¹<https://github.com/Jeamcarlosbraun/ppgca-upf-ds>