



Otimização da Qualidade do Leite

Uma Abordagem CRISP-DM

Objetivo do Projeto

A qualidade do leite é um fator **determinante para a rentabilidade** dos produtores e para a indústria de laticínios.

Objetivo Principal

Analisar e modelar os fatores que influenciam os níveis de **gordura e proteína** no leite, desenvolvendo modelos preditivos para otimizar esses componentes essenciais.

- ✓ Identificar padrões e correlações entre variáveis
- ✓ Desenvolver modelos preditivos para gordura e proteína
- ✓ Fornecer recomendações para otimização da qualidade



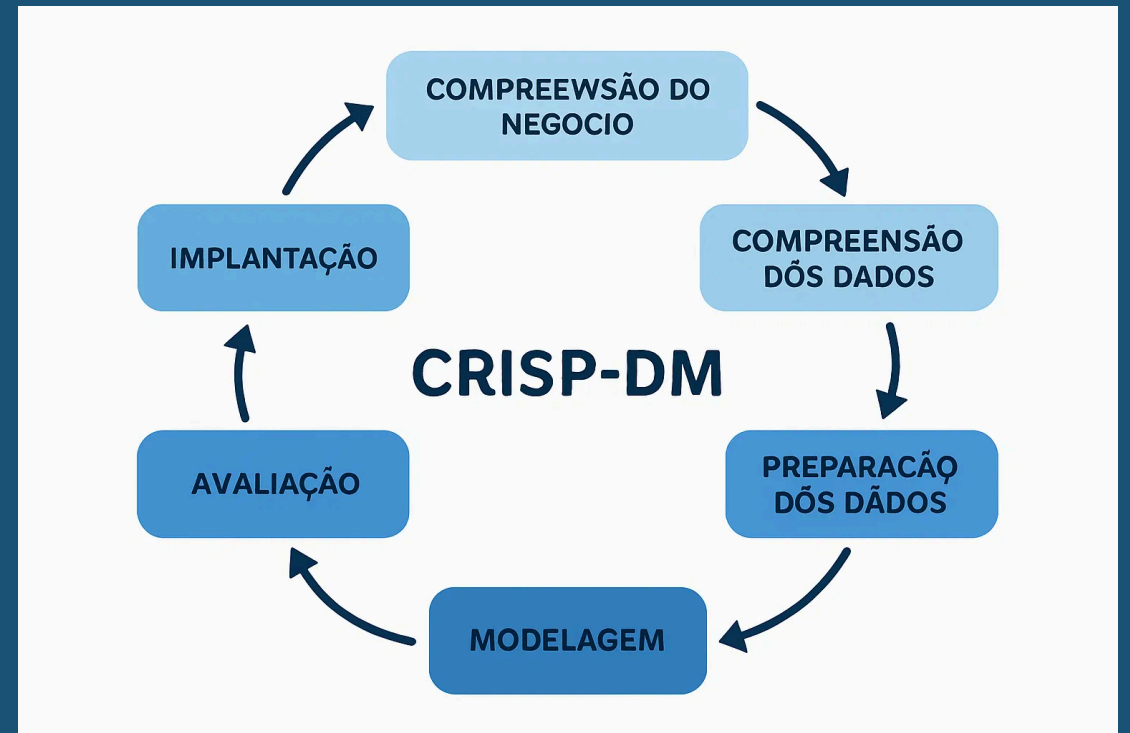
Dataset e Metodologia

Dataset:	milk_sessions.csv
Registros:	35.517
Variáveis:	44
Período:	Janeiro a Abril de 2020

Principais variáveis: DailyYield_KG, DailyFat_P, DailyProtein_P, LactationNumber, DIM (dias em lactação)

Transformações aplicadas:

- Conversão de datas para extração de features temporais
- Remoção de colunas com muitos valores nulos
- Normalização de features numéricas

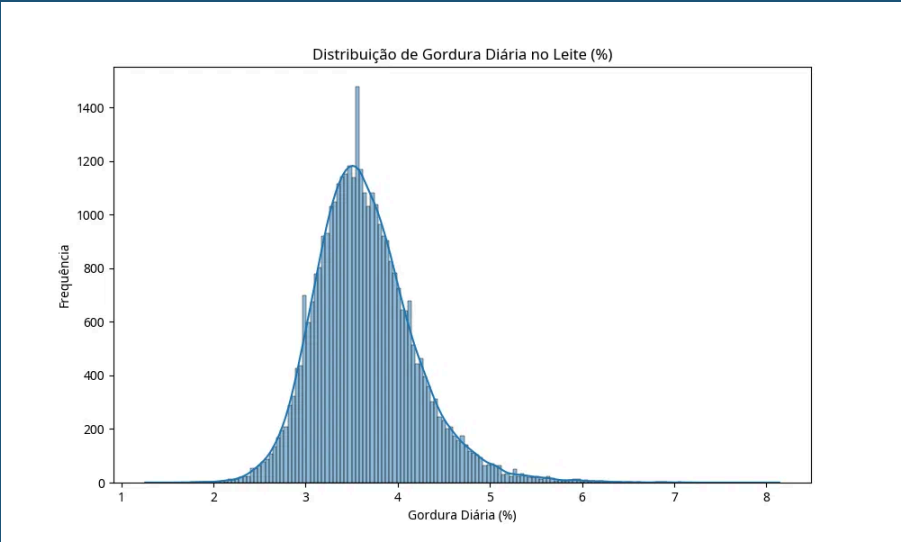


Análise Exploratória

Principais insights da análise exploratória:

- ✔ Gordura apresenta maior variabilidade (desvio padrão: 0,55%) do que proteína (desvio padrão: 0,24%)
- ✔ Correlação negativa fraca (-0,07) entre gordura e proteína diárias
- ✔ Forte correlação entre gordura diária e ácidos graxos (Mufa: 0,68, Pufa: 0,63)
- ✔ Número de lactações influencia significativamente os níveis de gordura e proteína

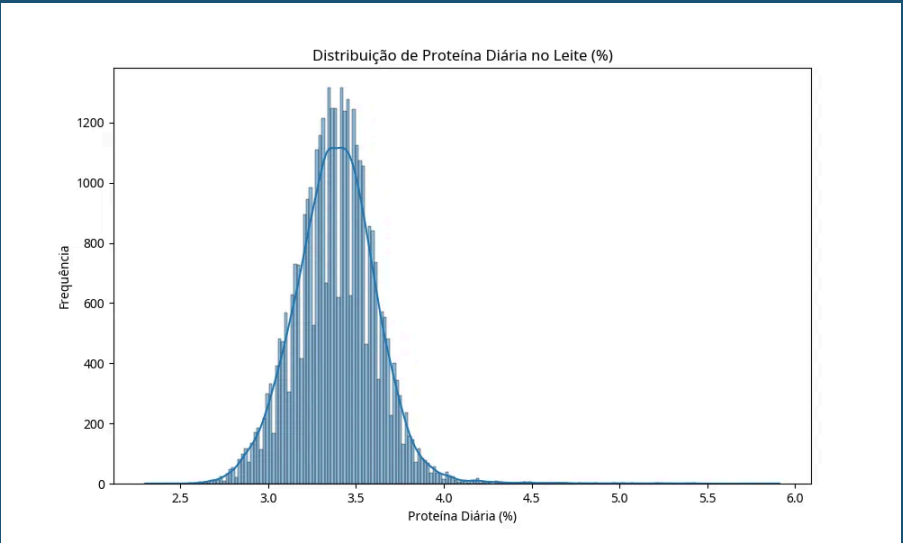
Distribuição do Percentual de Gordura Diária



Estatísticas descritivas:

Variável	Média	Mín-Máx
DailyFat_P	3,65%	1,26% - 8,14%
DailyProtein_P	3,39%	2,30% - 5,91%
DailyYield_KG	39,87	0,22 - 85,55

Distribuição do Percentual de Proteína Diária



Modelagem e Resultados

Utilizamos **Regressão Linear** para desenvolver modelos preditivos para os componentes de gordura e proteína no leite.

Modelo para Gordura Diária (DailyFat_P)

✓ Erro Quadrático Médio (MSE): 0,1057

☒ R-quadrado (R^2): 0,6548

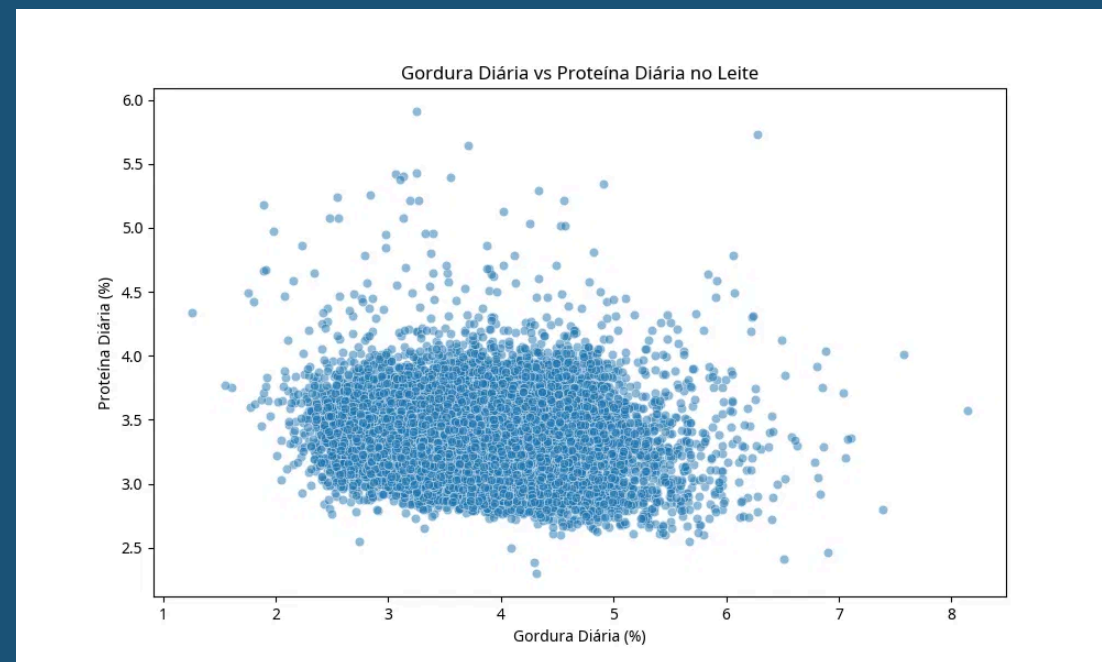
Modelo para Proteína Diária (DailyProtein_P)

✓ Erro Quadrático Médio (MSE): 0,0261

☒ R-quadrado (R^2): 0,5567

Interpretação: O modelo de gordura explica 65,48% da variância, enquanto o modelo de proteína explica 55,67%, indicando boa capacidade preditiva para ambos os componentes.

Features importantes: Número de lactações, dias em lactação (DIM), e composição de ácidos graxos.



Conclusões e Recomendações

- ✓ Os modelos preditivos alcançaram **bom desempenho**, com R^2 de 0,65 para gordura e 0,56 para proteína.
- ✓ O **número de lactações** tem impacto significativo nos níveis de gordura e proteína.
- ✓ Existe uma **correlação negativa fraca** entre gordura e proteína no leite.

Recomendações

- 💡 **Estratificar o rebanho** por número de lactações para manejo nutricional específico.
- 💡 **Implementar ajustes na dieta** baseados nos modelos preditivos.
- 💡 **Expandir o modelo** para incluir fatores ambientais e sazonais.

