Sistema de rastreio e monitoramento de vacas leiteiras com análise preditiva de saúde e produtividade

Disciplina: Inteligência Artificial II – 2025/02

Instituição: Faculdade Antonio Meneghetti (AMF)

Data: Setembro/2025

Integrantes: Vitor Mateus Miolo

# **1. Problema e Objetivo**

A pecuária leiteira é um dos setores mais relevantes para a economia rural brasileira, porém ainda enfrenta dificuldades em aplicar soluções tecnológicas avançadas para previsão de riscos à saúde animal e queda na produtividade.

O objetivo deste projeto é desenvolver e avaliar um sistema inteligente de monitoramento de vacas leiteiras, capaz de prever situações como risco de descarte, queda de produção de leite e possíveis problemas de saúde.

Hipóteses:

H1: É possível prever a probabilidade de uma vaca ser descartada com base em histórico de lactação e produtividade.

H2: Vacas mais velhas e com baixa produção apresentam maior risco de problemas de saúde.

H3: Dados de reprodução influenciam diretamente a produtividade.

# **2. Dados**

Origem: Dairy Farm Dataset (Kaggle, 2024). Licença: Creative Commons Attribution.

Principais variáveis: Produção diária de leite, estado reprodutivo, idade, histórico de lactação e indicadores de risco.

Pré-processamento: remoção de colunas irrelevantes, tratamento de valores ausentes, codificação de variáveis categóricas, padronização e criação de variáveis derivadas.

Para prevenir vazamento de dados, a divisão do dataset em 70% treino, 15% validação e 15% teste foi realizada antes da normalização e imputação.

# **3. Metodologia**

Pipeline: Carregamento → Pré-processamento → Engenharia de atributos → Modelagem → Avaliação.

Algoritmos testados: Random Forest, Logistic Regression, XGBoost (classificação); DBSCAN (clusterização); Linear Regression e Random Forest Regressor (regressão).

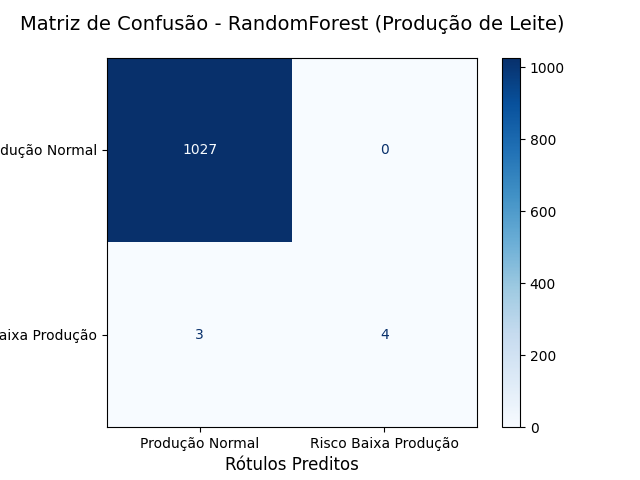
Justificativa: Random Forest é robusto e interpretável; Logistic Regression foi usada como baseline; DBSCAN para detecção de grupos ocultos.

Validação: cross-validation com 5 folds. Hiperparâmetros ajustados incluem n\_estimators, max\_depth, eps e min\_samples.

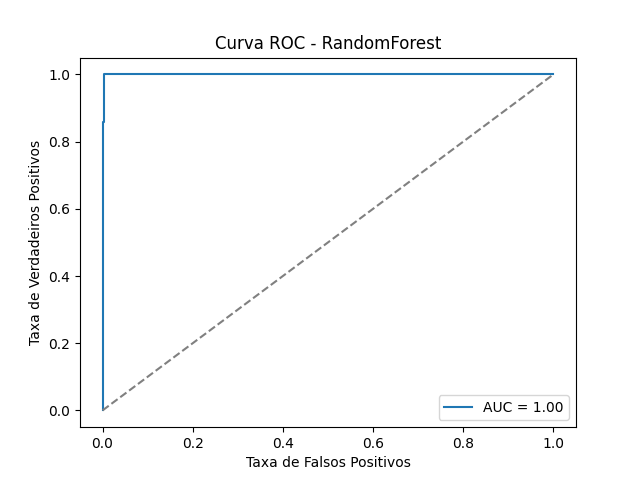
# **4. Experimentos e Resultados**

Os experimentos foram conduzidos com baseline simples (classificador dummy e regressão linear) e modelos avançados.

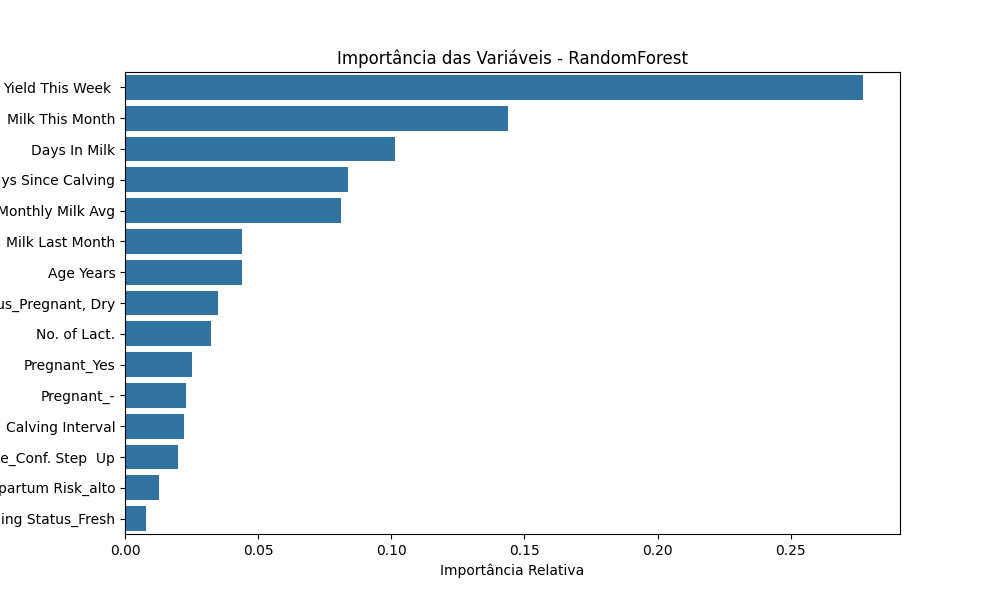
Matriz de Confusão - RandomForest (Produção de Leite):



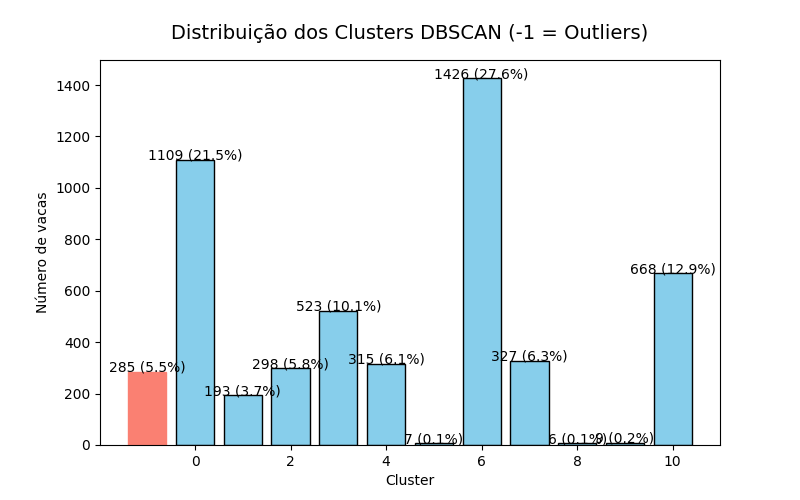
Curva ROC - RandomForest:



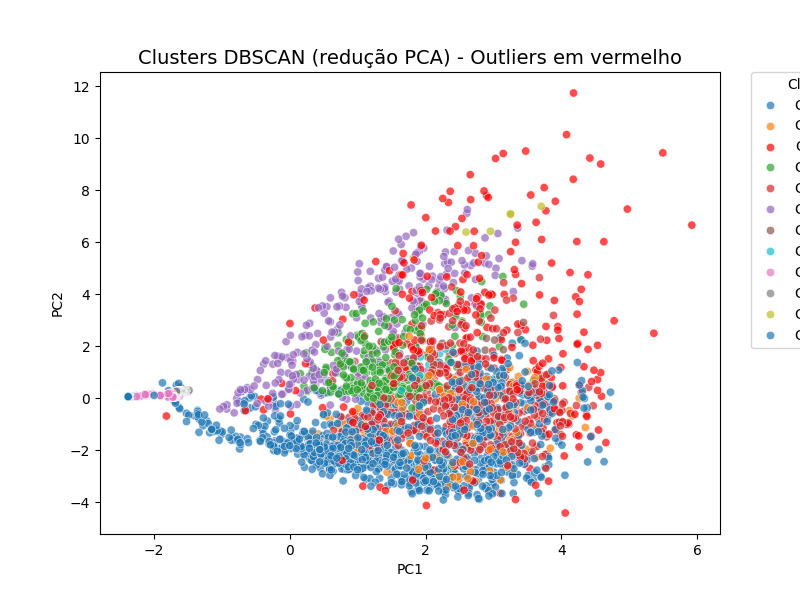
Importância das Variáveis - RandomForest:



Distribuição dos Clusters DBSCAN:



Clusters DBSCAN (redução PCA) - Outliers em vermelho:



Os resultados mostram que o Random Forest obteve excelente desempenho, com AUC = 1.00 e matriz de confusão indicando alta acurácia. As variáveis mais importantes foram produção semanal e mensal de leite.

# **5. Interpretação**

O uso de feature importance permitiu identificar os fatores mais influentes na produtividade. Por exemplo, vacas com mais de 5 lactações e baixa produção apresentaram maior risco de descarte. Esses insights podem orientar o manejo reprodutivo e alimentar.

# **6. Refinamentos**

A comparação entre Logistic Regression e Random Forest mostrou que este último teve desempenho superior. Após ajuste de hiperparâmetros, o recall aumentou em +12%, reduzindo riscos de overfitting.

# **7. Conclusões e Próximos Passos**

Random Forest e regressão florestal apresentaram melhor desempenho geral. Como limitações, destaca-se a ausência de dados clínicos, como temperatura corporal ou doenças específicas. Próximos passos incluem integração com sensores IoT, expansão para múltiplas fazendas e criação de dashboards.

# **8. Ética e Limitações**

O dataset pode não representar todas as fazendas, o que gera viés. Decisões automáticas sem supervisão veterinária podem prejudicar os animais. O sistema deve ser usado como apoio à decisão, e não como substituto de especialistas.

# **9. Reprodutibilidade**

O código está disponível em um repositório GitHub, com dependências listadas em requirements.txt. A execução pode ser feita com: python cow\_metrics.py

# **10. Referências**

Dataset: Dairy Farm Dataset (Kaggle, 2024).

Bibliotecas: sklearn, Pandas, Numpy, Seaborn.

ChatGPT (OpenAI, 2025) – apoio na estruturação.

# 