

Enterprise Challenge - Sprint 3 - Ingredion

Luana Porto Pereira Gomes

Luma Oliveira

Paulo Bernardes

Priscilla Garcia de Oliveira

Introdução

Este relatório apresenta o resultado final do desafio proposto ao longo das Sprints 1, 2 e 3 do Challenge Ingredion.

Sprint 1 – Coleta e Preparação dos Dados

Na primeira etapa, nosso grupo realizou a **pesquisa e coleta de dados históricos de produtividade agrícola** e imagens de NDVI, utilizando fontes públicas como IBGE, e SATVeg (INPE). Também foram feitos os primeiros tratamentos, organização das bases por município e safra, e análises exploratórias para entender a estrutura dos dados.

Sprint 2 – Desenvolvimento do Modelo Preditivo

Com os dados tratados, foi desenvolvida uma **primeira versão de modelo de IA** para prever produtividade a partir do NDVI. Foram testados diferentes algoritmos de regressão e classificação. Os resultados mostraram que o modelo ainda tinha espaço para melhorias, principalmente em termos de variáveis explicativas.

Sprint 3 – Validação com Dados Reais

Nesta última etapa, o modelo foi **validado com dados reais de produtividade**, comparando os valores previstos com os resultados oficiais obtidos nas bases públicas. Também foram incluídas **variáveis climáticas**, para aumentar a precisão da previsão e enriquecer a análise estatística. Esta Sprint representa a consolidação do projeto, unindo dados, modelo e avaliação crítica.

Este documento reúne toda a jornada do projeto, desde a coleta até a validação estatística e sugestões de melhorias para aplicações reais.

Validação do Modelo de IA com Dados Reais

Relatório Final

1. Metodologia de Coleta de Dados Históricos

Nesta etapa, foram utilizados dados publicos de NDVI e produtividade agricola real para validar o modelo preditivo criado na Sprint 2. As fontes consultadas foram:

- IBGE/SIDRA - Produtividade agricola (kg/ha) por municipio e safra;
- SATVeg (INPE) - Imagens de NDVI para os periodos correspondentes;
- INMET - Dados climaticos diarios agregados por safra: precipitacao, temperatura, umidade relativa e numero de dias com ou sem chuva.

Foram selecionados dados para 3 municipios do Mato Grosso (Campo Novo do Parecis, Nova Mutum e Sorriso), com abrangencia das safras de 2019 a 2023, focando na cultura da soja.

2. Leitura e Exploração dos Dados

A base utilizada contém as seguintes variáveis:

- Município e Safra
- Área plantada, área colhida, produção e rendimento (kg/ha)
- NDVI médio da safra
- Temperatura máxima e mínima, precipitação total e umidade relativa média
- Número de dias com chuva > 20mm e sem chuva > 1mm

Após a leitura, verificamos:

- Nenhuma coluna com valores ausentes.
- Tipos de dados coerentes para análise.

3. Análise estatística e Correlação

Correlação de Pearson e Spearman

Foram calculadas correlações entre as variáveis, com destaque para:

- NDVI e Produtividade: correlação fraca (Pearson = 0,1663; Spearman = 0,1259)
- NDVI e Temp. máxima: moderada (Pearson = 0,5885; Spearman = 0,6416)
- NDVI e Umidade: negativa (Pearson = -0,4743; Spearman = -0,3557)
- Produtividade e Temp. máxima: moderada (Pearson = 0,5508; Spearman = 0,4555)
- Produtividade e Umidade: negativa (Pearson = -0,5443; Spearman = -0,4122)

As correlações de Spearman, que analisam relações não lineares ou monotônicas, reforçam os padrões já identificados por Pearson, sugerindo que as variáveis climáticas têm impacto relevante na produtividade agrícola.

4. Regressão Linear Simples

Cada variável independente foi usada para prever a produtividade (kg/ha):

- NDVI médio da safra | $R^2 = 0,0276$ | Baixo poder preditivo
- Precipitação total | $R^2 = 0,0085$ | Relação praticamente inexistente
- Temp. máxima | $R^2 = 0,3033$ | Melhor desempenho individual
- Temp. mínima | $R^2 = 0,0012$ | Sem relação
- Umidade média | $R^2 = 0,2963$ | Relação negativa moderada

5. Regressão Linear Multipla (NDVI + Clima)

Utilizamos as variáveis:

- NDVI médio da safra
- Temperatura máxima
- Dias com chuva > 20mm
- Umidade média

O modelo múltiplo alcançou $R^2 = 0,412$ (41,2%), mostrando melhora considerável no poder preditivo.

Equação:

Produtividade = $-1605.24 \cdot \text{NDVI} + 91.14 \cdot \text{Temp_max} + -1.75 \cdot \text{Dias_chuva} + -43.71 \cdot \text{Umidade} + 5582.48$

6. Visualizações Geradas

- Gráficos de dispersão com linha de regressão (linear simples) .
- Gráfico de dispersão com previsão vs valor real (regressão múltipla).

Essas visualizações reforçam os resultados estatísticos e facilitam a interpretação visual.

7. Limitações da Análise

- Base com apenas 15 registros (3 municípios x 5 safras)
- Dados agregados por safra, sem granularidade mensal ou por talhão
- Não inclui fatores como solo, pragas, manejo ou uso de insumos

8. Sugestões de Melhoria

- Ampliar o número de registros e incluir novas regiões
- Considerar variáveis agronômicas e operacionais
- Testar modelos mais robustos (Ex: regressão não linear, árvores de decisão, redes neurais)
- Avaliar o NDVI em diferentes estágios da safra

9. Conclusão

O NDVI mostrou-se limitado como preditor isolado de produtividade. Entretanto, ao ser combinado com variáveis climáticas, o modelo explicou 41,2% da variação da produtividade.

A aplicação de modelos de IA com múltiplas variáveis mostra-se promissora para estimar produtividade em larga escala, desde que sejam considerados mais dados e fatores relevantes do contexto agrícola.

10. Referencias das Bases Publicas Utilizadas

- IBGE - SIDRA: <https://sidra.ibge.gov.br>
- INMET - Instituto Nacional de Meteorologia: <https://bdmep.inmet.gov.br>
- SATVeg (INPE): <http://www.obt.inpe.br/satveg/>