

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. Técnicas Estatísticas Aplicadas

Correlação:

- Pearson: $R = 0.166$ (fraca)
- Spearman: $R = 0.126$ (fraca)
- Informacao Mutua: 0.009 (dependencia muito baixa)

Regressões:

- Linear simples (NDVI): $R^2 = 0.028$
- Linear multipla (NDVI + clima): $R^2 = 0.412$

3. Analise dos Graficos Gerados

- Grafico de dispersao mostrou relacao fraca entre NDVI e produtividade.

- Regressao linear simples indicou baixa inclinacao da reta.
- Grafico por municipio/safra revelou variacoes relevantes entre os anos.
- Correlacao com clima:
 - Temperatura maxima media: $r = 0.55$ (positiva)
 - Umidade media: $r = -0.54$ (negativa)
 - Dias com chuva $> 20\text{mm}$: $r = 0.36$ (positiva)

4. Discussão Crítica e Melhorias

O NDVI isoladamente mostrou-se fraco para prever produtividade, mas combinado com variáveis climáticas apresentou um modelo com maior poder explicativo ($R^2 = 0.412$).

Fatores que podem ter influenciado os resultados:

- Qualidade das imagens NDVI
- Pragas ou doenças não mapeadas
- Eventos climáticos extremos

Melhorias sugeridas:

- Incluir mais variáveis (tipo de solo, insumos)
- Usar imagens NDVI com maior frequência
- Testar modelos não lineares (Random Forest, Redes Neurais)
- Expandir a análise para mais municípios

5. 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/>