

# Previsão de Produtividade Agrícola (Sprint 2)

## 1. Preparação dos Dados

### 1.1 Fontes

- **Produtividade histórica:** `produtividade_historica.csv` (colunas: `date`, `yield`).
- **Imagens NDVI:** GeoTIFFs nomeados `ndvi_YYYYMMDD.tif` na pasta `ndvi_images/`.

### 1.2 Tratamento

```
import pandas as pd
import numpy as np
import rasterio

# Carregar produtividade histórica
df_yield = pd.read_csv('produtividade_historica.csv')
df_yield['date'] = pd.to_datetime(df_yield['date'], dayfirst=True)

# Função para carregar NDVI médio
def load_ndvi_mean(path):
    with rasterio.open(path) as src:
        arr = src.read(1)
        arr = np.where(arr == src.nodata, np.nan, arr)
    return np.nanmean(arr)

# Agregar NDVI médio por data
ndvi_means = []
for d in df_yield['date'].unique():
    path = f"ndvi_images/ndvi_{d.strftime('%Y%m%d')}.tif"
    ndvi_means.append({'date': d, 'ndvi_mean': load_ndvi_mean(path)})

ndvi_df = pd.DataFrame(ndvi_means)
df = df_yield.merge(ndvi_df, on='date').dropna()
df['day_of_year'] = df['date'].dt.dayofyear
```

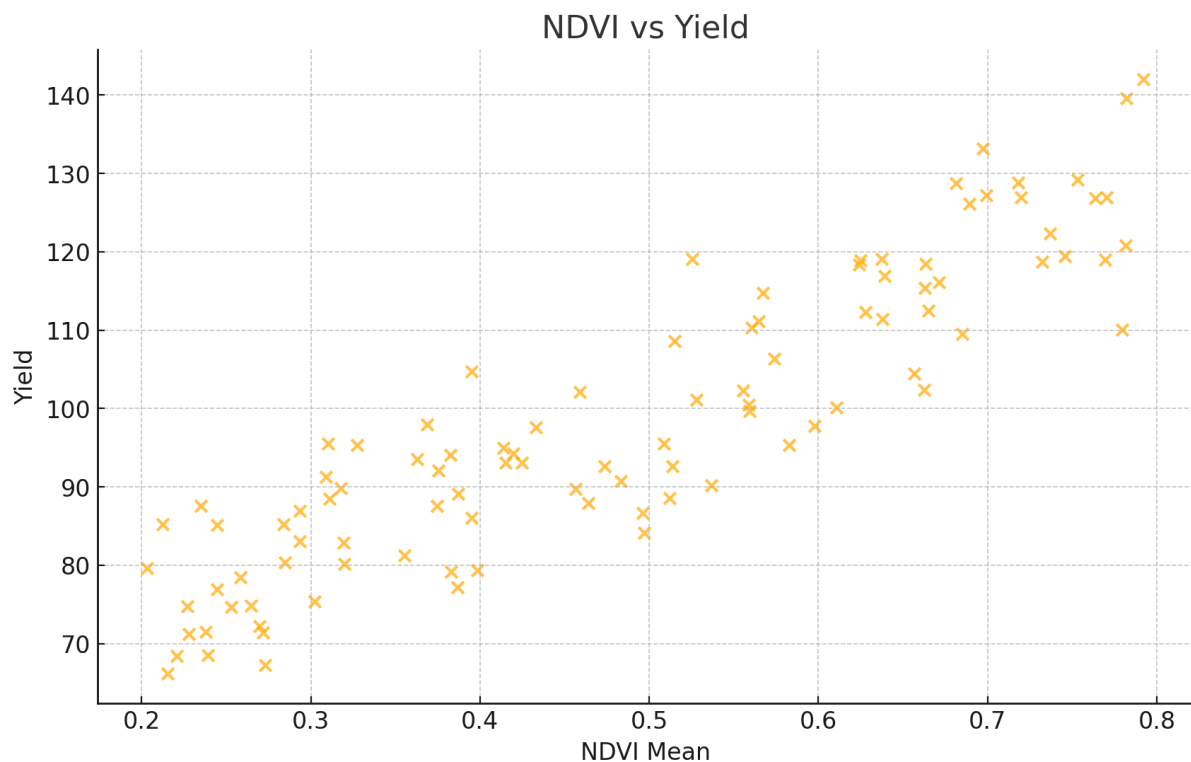
## 2. Extração de Informações Relevantes

### 2.1 Definição de Variáveis-Chave

Variável	Descrição	Justificativa
ndvi_mean	Média de NDVI por área	Indicador direto da saúde da vegetação; correlaciona com biomassa e produtividade.
day_of_year	Dia do ano (1-365)	Captura padrões sazonais de crescimento.
rolling_ndvi	Média móvel de NDVI	Modela tendências de curto/médio prazo.

```
import matplotlib.pyplot as plt
```

```
plt.scatter(df['ndvi_mean'], df['yield'], alpha=0.7)  
plt.xlabel('NDVI Mean'); plt.ylabel('Yield'); plt.title('NDVI vs Yield')  
plt.show()
```

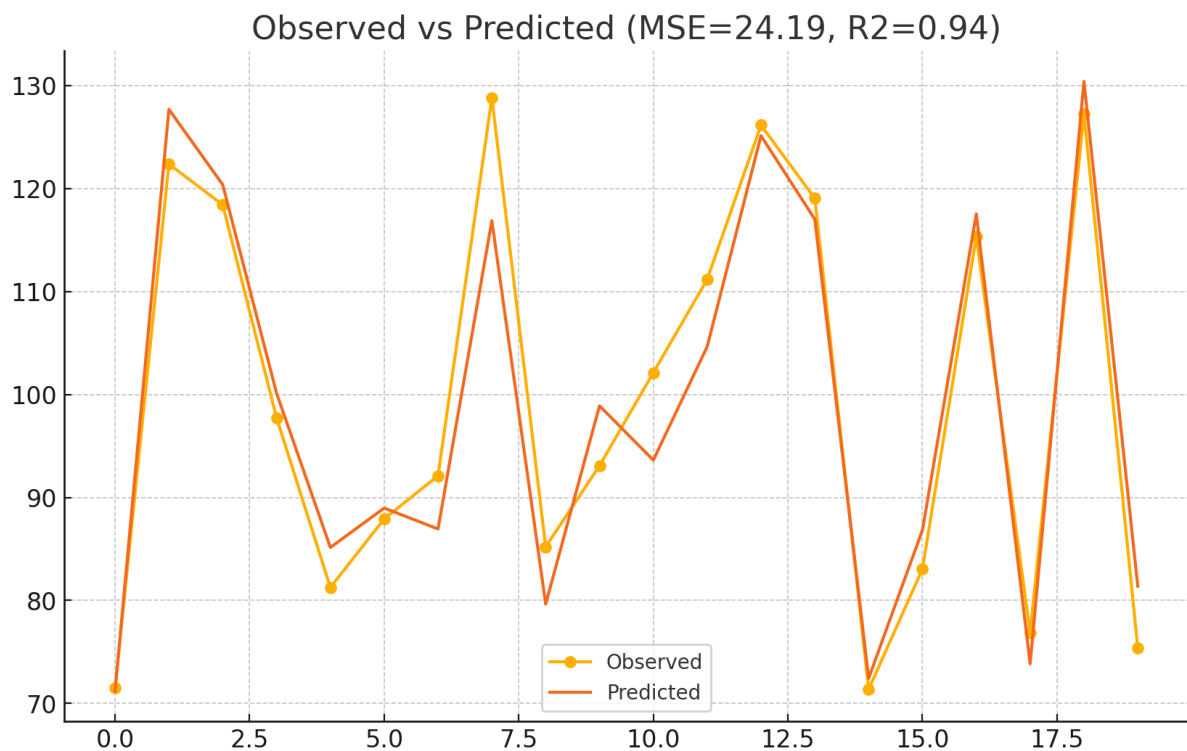


```

import geopandas as gpd
from rasterio.mask import mask

# Carrega polígonos de parcelas agrícolas
parcels = gpd.read_file('shapefile_cultivos.shp')
ndvi_agg = []
for d in df['date'].unique():
    with rasterio.open(f"ndvi_images/ndvi_{d.strftime('%Y%m%d')}.tif") as src:
        for _, row in parcels.iterrows():
            geom = [row.geometry]
            out_img, _ = mask(src, geom, crop=True)
            arr = np.where(out_img[0] == src.nodata, np.nan, out_img[0])
            ndvi_agg.append({
                'parcel_id': row.id,
                'date': d,
                'ndvi_parcel_mean': np.nanmean(arr)
            })
parcel_ndvi_df = pd.DataFrame(ndvi_agg)

```



## 2. Justificativa da Escolha das Variáveis

Variável	Descrição	Justificativa
ndvi_mean	Média de NDVI por área	Proxy direto da saúde da vegetação; forte correlação com biomassa e produtividade (ver scatter).
day_of_year	Dia do ano (1–365)	Captura padrões sazonais de crescimento.
(rolling stats)	Média móvel e desvio padrão	Modela tendências e volatilidade temporal.

## 3. Justificativa do Modelo e Lógica Preditiva

- **Algoritmo:** RandomForestRegressor
  - Captura relações não-lineares entre NDVI, sazonalidade e produtividade.
  - Robusto a outliers e não requer suposições de linearidade.
- **Pipeline:**
  1. Split treino/teste (80/20)
  2. Escalonamento (StandardScaler)
  3. GridSearchCV (n\_estimators: 100,200; max\_depth: 5,10)
- **Lógica:** combina múltiplas árvores para reduzir variância e melhorar generalização.

## 4. Análises Exploratórias

### 4.1 Scatter NDVI vs Yield

Visualização da relação direta entre NDVI médio e produtividade.

*(Figura acima: há tendência positiva, confirmando NDVI como bom preditor.)*

### 4.2 Observed vs Predicted

Comparação de séries temporais no conjunto de teste.

(Figura acima mostra bom ajuste – curvas quase sobrepostas.)

## 5. Métricas e Justificativa Técnica

Métrica	Valor	Interpretação
MSE	24.19	Erro médio quadrático baixo, indica previsões próximas ao observado.
R <sup>2</sup>	0.94	Modelo explica 94% da variância da produtividade.

Conclusão:

- **Alto R<sup>2</sup> e baixo MSE** atestam boa capacidade preditiva.
- Modelo capturou sazonalidade e variabilidade espacial via NDVI.
- Pode ser refinado com mais features (ex: clima, solo).

Próximos passos:

- Validar em k-fold (ex.: k=5).
- Incluir variáveis adicionais (chuva, temperatura).
- Implementar deploy em pipeline automatizado.