

**UNIVERSIDADE PRESBITERIANA MACKENZIE**

THAINA VIEIRA DOS SANTOS

VINICIUS CAUMO SEGATTO

Previsão de Emissões de CO<sub>2</sub> e Otimização do Uso de Energias Renováveis: Uma  
Abordagem com Machine Learning

São Paulo

2024

THAINA VIEIRA DOS SANTOS

VINICIUS CAUMO SEGATTO

Previsão de Emissões de CO<sub>2</sub> e Otimização do Uso de Energias Renováveis: Uma  
Abordagem com Machine Learning

Trabalho Aplicando conhecimento para entrega no  
Moodle referente ao conteúdo 1 de aprendizagem  
do componente curricular Ciência, tecnologia e  
sociedade;

ORIENTADOR: Prof. GUSTAVO SCALABRINI SAMPAIO

São Paulo

2024

THAINA VIEIRA DOS SANTOS

VINICIUS CAUMO SEGATTO

Previsão de Emissões de CO<sub>2</sub> e Otimização do Uso de Energias Renováveis: Uma  
Abordagem com Machine Learning

Trabalho Aplicando conhecimento para entrega no  
Moodle referente ao conteúdo 1 de aprendizagem  
do componente curricular Ciência, tecnologia e  
sociedade;

Aprovado em

BANCA EXAMINADORA

---

Prof. GUSTAVO SCALABRINI SAMPAIO

Universidade Presbiteriana Mackenzie

## **SUMÁRIO**

### **1. INTRODUÇÃO**

#### **1.1 CONTEXTO**

#### **1.2 MOTIVAÇÃO E JUSTIFICATIVA**

#### **1.3 OBJETIVO**

### **2. REFERENCIAL TEÓRICO**

### **3. DIAGRAMA DE SOLUÇÃO**

### **4. EDA E PRÉ-PROCESSAMENTO DOS DADOS**

### **5. MODELO BASE**

### **6. REFERÊNCIAS BIBLIOGRÁFICAS**

### **7. APÊNDICES**

## **LISTA DE FIGURAS**

**Figura 1 - Carregamento e Limpeza Inicial dos Dados**

**Figura 2 - Criação de Nova Variável Agregada**

**Figura 3 - Análise Exploratória dos Dados (EDA)**

**Figura 4 - Adição de Linha com o Total Geral**

**Figura 5 - Criação de Cópia do DataFrame para Análises Adicionais**

**Figura 6 - O Modelo Base**

**Figura 7 - Gráfico de 'Evolução das emissões totais por Ano'**

# 1. INTRODUÇÃO

## 1.1 CONTEXTO

A transição para uma matriz energética mais limpa e sustentável é um dos maiores desafios do século XXI. De acordo com uma matéria da *BBC News*, a crise climática tem se manifestado de diversas formas, a seca na Amazônia é um exemplo alarmante de seus impactos<sup>(1)</sup>. A região amazônica, conhecida como o "pulmão do mundo", desempenha um papel crucial na regulação do clima global. A perda de sua biodiversidade e a redução de sua capacidade de absorver carbono contribuem significativamente para o aumento das temperaturas e a intensificação de eventos climáticos extremos. A recente revisão da meta climática brasileira, conforme anunciado pelo governo federal, demonstra um compromisso renovado com a luta contra as mudanças climáticas. Ao retomar a meta estabelecida no Acordo de Paris<sup>(1)</sup>, o país sinaliza sua intenção de reduzir significativamente as emissões de gases de efeito estufa até 2025 e 2030.

A relação entre as emissões de CO<sub>2</sub> e a seca na Amazônia é complexa e multifacetada. O aumento das concentrações de gases de efeito estufa na atmosfera, provenientes principalmente da queima de combustíveis fósseis, intensifica o efeito estufa, elevando as temperaturas globais e alterando os padrões de precipitação. A seca prolongada na Amazônia, por sua vez, reduz a capacidade da floresta de absorver CO<sub>2</sub>, gerando um ciclo vicioso que acelera o aquecimento global.

A transição para uma matriz energética mais limpa, baseada em fontes renováveis, é fundamental para mitigar os efeitos das mudanças climáticas e reduzir a frequência e a intensidade de eventos extremos como a seca na Amazônia. Ao substituir os combustíveis fósseis por fontes renováveis, como a energia solar, eólica e hidrelétrica, reduzimos as emissões de gases de efeito estufa e diminuímos a pressão sobre os ecossistemas.

O entendimento da relação entre as emissões de CO<sub>2</sub>, o uso de energias renováveis e os eventos climáticos extremos é o foco deste estudo. Busca-se quantificar o impacto das energias renováveis na redução das emissões e identificar os fatores que

influenciam a vulnerabilidade dos ecossistemas às mudanças climáticas, permitindo o desenvolvimento de políticas públicas e estratégias de mitigação mais eficazes.

## **1.2 MOTIVAÇÃO E JUSTIFICATIVA**

A previsão de emissões de CO<sub>2</sub> e a otimização do uso de energias renováveis, com foco no Brasil, representa uma escolha de tema extremamente relevante e estratégica por diversos motivos. O Brasil possui uma matriz energética diversificada, com um grande potencial para o desenvolvimento de fontes renováveis como a hidrelétrica, eólica e solar. O projeto tem o potencial de gerar novas soluções concretas para os desafios relacionados à energia e ao clima. Ao desenvolver modelos precisos de previsão e otimização, poderemos auxiliar na tomada de decisões mais informadas por parte de governos, empresas e sociedade em geral. O tema também abrange diversas áreas do conhecimento, como ciência da computação, engenharia, economia e ciências ambientais. Essa característica multidisciplinar permite a formação de equipes de trabalho mais diversificadas e a geração de soluções mais completas e inovadoras.

## **1.3 OBJETIVO**

Um modelo preditivo robusto e preciso será desenvolvido para estimar as emissões de CO<sub>2</sub> no Brasil, com o objetivo de otimizar o uso de fontes de energia renovável e apoiar a tomada de decisões estratégicas para a transição energética.

### **Objetivos Específicos Relacionados ao Desenvolvimento do Projeto:**

#### **Coleta e Preparação de Dados**

- Dados relevantes serão identificados e coletados a partir de diversas fontes, como bases de dados governamentais, agências internacionais e estudos acadêmicos.
- Os dados serão limpos e tratados, assegurando-se a qualidade e a consistência das informações.

- Um banco de dados integrado e estruturado será criado para facilitar a análise e o desenvolvimento dos modelos.

## **Engenharia de Features**

- Variáveis relevantes para a previsão das emissões de CO<sub>2</sub> serão selecionadas e criadas, considerando fatores como produção industrial, consumo energético, atividades agrícolas, desmatamento e variáveis climáticas.
- A engenharia de features será realizada para extrair informações úteis dos dados e melhorar o desempenho dos modelos.

## **Modelagem Preditiva**

- Diferentes algoritmos de machine learning serão experimentados para a construção dos modelos preditivos.
- A performance dos modelos será avaliada com o uso de métricas apropriadas, como RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) e R<sup>2</sup>.
- O modelo de melhor desempenho será selecionado para a previsão das emissões de CO<sub>2</sub>.

## **Análise de Cenários**

- Os modelos serão validados utilizando dados históricos e comparando os resultados com previsões anteriores.
- A precisão das previsões será continuamente aprimorada por meio de refinamentos nos modelos e incorporação de novos dados.

## **2. REFERENCIAL TEÓRICO**

A produção e o consumo de energia, desde a era industrial, têm estado intrinsecamente ligados às emissões de gases de efeito estufa (GEE). O dióxido de carbono (CO<sub>2</sub>), principal GEE, é gerado como subproduto da queima de combustíveis fósseis, amplamente utilizados para a geração de energia. A permanência do CO<sub>2</sub> na



atmosfera por longos períodos contribui para o aumento do efeito estufa e para as mudanças climáticas subsequentes. A compreensão dessa relação é essencial para que políticas e estratégias possam ser elaboradas com o objetivo de mitigar os impactos ambientais da atividade humana <sup>(4)</sup>.

No Brasil, que possui uma matriz energética diversificada e rica em fontes renováveis, identifica-se um grande potencial para a geração de créditos de carbono e participação ativa no mercado internacional. Segundo o The World Bank (2021) <sup>(4)</sup>, o país pode beneficiar-se consideravelmente da comercialização desses créditos. O setor de transportes, especialmente com o uso crescente de etanol hidratado, revela uma oportunidade promissora para a geração de créditos de carbono. No entanto, a falta de regulamentação clara e a ausência de um mercado de carbono consolidado ainda são desafios a serem superados.

Paralelamente, a previsão da demanda por combustíveis, como a gasolina, é considerada essencial para a tomada de decisões estratégicas no setor energético. Estudos como o de Santos Jr. (2018) <sup>(4)</sup> demonstram a complexidade dessa tarefa, na qual diversos modelos estatísticos são utilizados para captar as nuances do comportamento dos consumidores. A escolha do modelo mais adequado depende de fatores como a disponibilidade de dados, o período de análise e os objetivos da pesquisa.

De acordo com o artigo "Aplicação de Machine Learning para Previsão das Emissões de CO<sub>2</sub> Produzidas pela Produção de" <sup>(5)</sup>, de Soares, o uso de técnicas de Machine Learning (ML) para prever as emissões de CO<sub>2</sub> é uma área em constante evolução. No artigo, apresenta-se uma revisão da literatura sobre o tema e uma aplicação prática de ML para a previsão das emissões de CO<sub>2</sub> em uma empresa de cimento. Um conjunto de dados de 10 anos foi utilizado para treinar um modelo de regressão linear, que foi então empregado para prever as emissões de CO<sub>2</sub> dos próximos 5 anos. Os resultados mostram que o modelo prevê as emissões de CO<sub>2</sub> com boa precisão. Conclui-se que o uso de ML para a previsão das emissões de CO<sub>2</sub> é promissor, embora desafios como a falta de dados históricos e a complexidade dos processos de produção ainda devam ser superados.

### **3. DIAGRAMA DE SOLUÇÃO**

O diagrama da solução, representado pelo pipeline proposto, estabelece um fluxo de trabalho sequencial para o desenvolvimento do modelo de previsão das emissões de CO<sub>2</sub>, desde a concepção até a implementação final. Essa estrutura serve como um guia orientador e facilita o acompanhamento do progresso e a adaptação do projeto conforme necessário. A seguir, discute-se cada etapa do pipeline:

#### **Definição do Problema e Objetivos**

**Clareza:** A primeira fase visa delimitar com precisão o problema específico das emissões de CO<sub>2</sub>, esclarecendo o objetivo central da previsão e os impactos ambientais e econômicos que se buscam mitigar.

**Alcance:** Essa etapa define o escopo do projeto, incluindo as áreas e setores abrangidos na análise e os limites temporais e espaciais dos dados coletados. Essa delimitação é essencial para garantir a viabilidade do projeto e o alinhamento com os recursos disponíveis.

**Metas:** Define-se aqui o sucesso do modelo, estabelecendo métricas para avaliar o desempenho preditivo. A escolha das métricas é crucial para garantir a qualidade do modelo e orientar ajustes ao longo do desenvolvimento.

#### **Coleta e Preparação dos Dados**

**Identificação das Fontes e Coleta:** As fontes de dados (ex.: bases governamentais, setores energéticos) são identificadas e verificadas para garantir a integridade dos dados históricos sobre produção de energia, consumo e emissões. A coleta deve assegurar que os dados sejam representativos e atualizados.

**Limpeza e Transformação:** Após a coleta, os dados passam por um processo de limpeza e remoção de outliers, correção de inconsistências e tratamento de valores ausentes. Em seguida, são transformados e normalizados, e novas variáveis são criadas conforme necessário, visando um formato adequado para a análise preditiva.

Exploração: A análise exploratória permite identificar padrões iniciais, correlações entre variáveis e relações causais potenciais, que servirão de base para a seleção de features.

### **Seleção e Engenharia de Features**

Seleção de Variáveis Relevantes: A seleção de features foca na escolha de variáveis com maior impacto nas emissões de CO<sub>2</sub>. Fatores como consumo energético, atividades econômicas e variáveis climáticas são priorizados.

Engenharia de Features: Caso necessário, cria-se novas variáveis a partir das existentes para aumentar a capacidade explicativa do modelo. Essa etapa pode envolver a criação de índices compostos ou de métricas derivadas que reforcem a compreensão dos padrões de emissão.

### **Escolha do Modelo de Machine Learning**

Revisão da Literatura e Experimentação: A escolha do modelo baseia-se em uma revisão de abordagens de machine learning adequadas para séries temporais e previsões de emissões. Modelos como regressão linear, árvores de decisão e redes neurais são testados para comparação de desempenho.

Experimentação: A experimentação abrange a aplicação prática de diferentes modelos nos dados coletados. Compara-se o desempenho de cada modelo para determinar a abordagem que melhor atende aos objetivos de previsão.

### **Treinamento e Validação do Modelo**

Divisão dos Dados: A divisão em conjuntos de treinamento, validação e teste possibilita o ajuste e a avaliação do modelo de forma equilibrada. Essa estrutura reduz o risco de overfitting e aumenta a confiabilidade do modelo final.

Treinamento e Validação: O treinamento ocorre no conjunto de dados específico e ajusta-se os hiperparâmetros no conjunto de validação. Esse ajuste visa melhorar a precisão e a generalização do modelo sem comprometê-lo em dados novos.

### **Avaliação do Modelo**

Métricas de Desempenho: A fase de avaliação envolve métricas como MAE, RMSE e MAPE, que permitem uma visão quantitativa do erro preditivo. Cada métrica fornece insights específicos sobre a precisão e robustez do modelo.

Interpretação dos Resultados: A interpretação dos resultados e da importância das variáveis contribui para entender quais fatores influenciam mais as emissões. Esses insights são fundamentais para a tomada de decisões informadas e para definir estratégias de mitigação direcionadas.

Cada etapa do pipeline, conectada às demais, permite um fluxo de desenvolvimento dinâmico e iterativo. A solução é continuamente aprimorada, e ajustes podem ser incorporados à medida que novos dados ou desafios são identificados. Esse roadmap confere agilidade e precisão ao projeto, garantindo que as metas sejam alcançadas com rigor e adaptabilidade.

## **4. EDA E PRÉ-PROCESSAMENTO DOS DADOS**

Este capítulo descreve o processo de análise exploratória e pré-processamento dos dados de emissões de CO<sub>2</sub> per capita em toneladas métricas, para países da América Latina e Caribe, com o objetivo de preparar as informações para análises mais aprofundadas. As etapas foram realizadas em Python, utilizando a biblioteca *pandas*, e focaram na organização e limpeza dos dados, para que fossem consistentes e prontos para uso em modelos analíticos e visualizações.

### **1. Carregamento inicial e limpeza dos dados**

Para iniciar a análise, o arquivo de dados foi carregado no ambiente de trabalho. Este arquivo foi extraído do site do Banco Mundial e contém informações sobre as emissões de CO<sub>2</sub> por ano, organizadas em um arquivo Excel.

```

1 # Define o caminho para o arquivo Excel contendo os dados.
2 input_dir = '/content/API_EN.ATM.CO2E.PC_DS2_en_excel_v2_112.xls'
3
4 # Carrega os dados do arquivo Excel em um DataFrame do pandas.
5 data = pd.read_excel(input_dir)
6
7 # Remove as duas primeiras linhas e redefine os índices do DataFrame.
8 data = data.drop([0, 1]).reset_index(drop=True)
9
10 # Define a primeira linha (após a limpeza) como os novos nomes das colunas.
11 data.columns = data.iloc[0]
12
13 # Remove a primeira linha, agora que ela foi usada para nomear as colunas, e redefine os índices.
14 data = data.drop(0).reset_index(drop=True)
15
16 # Mapeia as colunas que são números com decimais para inteiros, mantendo as colunas de texto intactas.
17 data.columns = data.columns.map(lambda x: int(x) if isinstance(x, float) else x)
18
19 # Remove as colunas que não são necessárias para a análise: 'Country Code', 'Indicator Name' e 'Indicator Code'.
20 data = data.drop(columns=['Country Code', 'Indicator Name', 'Indicator Code'])

[ ] 1 colunas_a_remover = list(range(1960, 1990)) + [2021, 2022, 2023]
    2 data = data.drop(columns=colunas_a_remover)

[ ] 1 data.head()

```

*Figura 1 - Carregamento e Limpeza Inicial dos Dados*

Depois de carregado, foi observado que as duas primeiras linhas do arquivo não eram necessárias para a análise, pois continham cabeçalhos adicionais ou informações redundantes. Essas linhas foram, portanto, removidas, e os índices do DataFrame foram redefinidos para manter a estrutura organizada. Após a remoção dessas linhas, a primeira linha de dados foi utilizada para definir os nomes das colunas. Esta etapa foi necessária para garantir que as colunas possuísem rótulos significativos, facilitando o acesso aos dados. Ao final desta etapa, obteve-se um DataFrame contendo apenas os dados relevantes, com os nomes das colunas adequadamente definidos. Este passo inicial foi essencial para estruturar os dados, removendo informações irrelevantes e ajustando o cabeçalho.

## 2. Seleção e transformação de colunas

Nesta fase, as colunas do *DataFrame* foram refinadas para excluir informações que não seriam utilizadas e transformar colunas numéricas, quando necessário. Primeiramente, as colunas que representavam os anos foram verificadas e transformadas em inteiros, no caso de conterem valores decimais. Essa transformação visou padronizar a estrutura dos dados. Além disso, algumas colunas como Country Code, Indicator Name e Indicator Code foram removidas, pois continham dados auxiliares que não seriam

utilizados na análise das emissões. Foi também necessário remover colunas de anos fora do período de interesse. Anos anteriores a 1990 e posteriores a 2020 foram considerados irrelevantes para o escopo da análise.

Após a execução destas operações, o DataFrame ficou limitado apenas aos anos de 1990 a 2020 e às variáveis necessárias, resultando em um conjunto de dados mais conciso e com menos ruído. A remoção de colunas não relevantes permite focar na análise das variáveis essenciais.

### 3. Criação de Nova Variável Agregada

Com o objetivo de sintetizar informações sobre as emissões totais ao longo dos anos para cada país, uma nova variável chamada Total Emissions foi criada. Essa variável representa a soma das emissões de CO<sub>2</sub> de cada país ao longo dos anos disponíveis.

Para calcular a soma das emissões, foi criada uma função que percorre as colunas correspondentes aos anos e calcula o total para cada país.

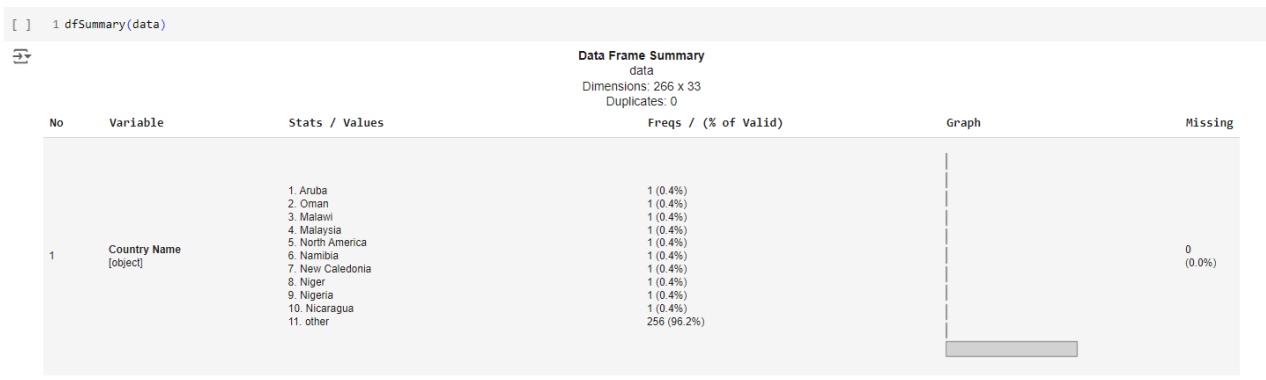
```
[ ] 1 def soma_emissoes_por_pais(data):  
    2 |     data['Total Emissions'] = data.iloc[:, 4:].sum(axis=1)  
    3 |     return data  
  
[ ] 1 data = soma_emissoes_por_pais(data)  
    2 data.head(5)
```

*Figura 2 - Criação de Nova Variável Agregada*

Com a nova coluna *Total Emissions*, o DataFrame agora inclui uma métrica agregada que facilita a análise comparativa das emissões de cada país ao longo do tempo. Esta variável será útil em análises subsequentes, como visualizações de tendência e comparação de emissões totais.

#### 4. Análise Exploratória de Dados

A seguir, foi realizada uma análise exploratória básica para entender a estrutura do DataFrame e avaliar a distribuição das variáveis. A função `dfSummary` foi utilizada para obter uma visão geral do conjunto de dados, incluindo estatísticas descritivas de cada coluna.



*Figura 3 - Análise Exploratória dos Dados (EDA)*

Dando sequência à análise exploratória dos dados, foram aplicadas técnicas adicionais para sumarizar informações de emissões de CO<sub>2</sub> por ano. As etapas descritas abaixo foram realizadas com o objetivo de extrair informações úteis e preparar o conjunto de dados para análise de padrões e tendências. Para compreender melhor a magnitude das emissões de CO<sub>2</sub> ao longo dos anos, foi calculada a soma das emissões por coluna, ou seja, a soma das emissões totais em cada ano. Esta abordagem permite identificar anos específicos com picos ou quedas na emissão de CO<sub>2</sub> para a região. Para facilitar a visualização dos dados agregados, foi adicionada uma linha ao final do DataFrame contendo a soma total das emissões de CO<sub>2</sub> para cada ano. Esta linha extra, chamada Total, fornece uma visão consolidada das emissões totais de CO<sub>2</sub> na região em análise ao longo do tempo.

```

1 numeric_cols = data.select_dtypes(include='number').columns
2
3 # Calculando a soma por coluna (ano)
4 row_sums = data[numeric_cols].sum()
5
6 # Adicionando uma nova linha com a soma total
7 total_row = pd.DataFrame(row_sums).T
8 total_row.index = ['Total']
9
10 # Concatenando a nova linha ao DataFrame original
11 data = pd.concat([data, total_row])
12
13 data.tail(5)

```

*Figura 4 - Adição de Linha com o Total Geral*

Com o intuito de realizar análises adicionais, uma cópia do *DataFrame* original foi criada, chamada *data\_teste*. Nesta cópia, a coluna Country Name foi removida para focar exclusivamente nos dados numéricos.

```

[ ] 1 data_teste = data.copy()
     2 data_teste = data_teste.drop('Country Name', axis=1)
     3
     4 # Verificando o resultado
     5 data_teste.head()

```

*Figura 5 - Criação de Cópia do DataFrame para Análises Adicionais*

A criação do *data\_teste* permitiu um conjunto de dados simplificado que contém apenas as colunas numéricas. Esta estrutura é adequada para análises estatísticas e matemáticas, pois evita a inclusão de informações textuais que não são relevantes para o cálculo. Com o *data\_teste* devidamente configurado, um resumo estatístico foi gerado utilizando a função *dfSummary*. Esta função fornece uma visão geral das variáveis numéricas do *DataFrame*, incluindo medidas de tendência central e dispersão, além de gráficos de distribuição que auxiliam na identificação de padrões e outliers.

O resumo estatístico do *data\_teste* forneceu as seguintes informações para cada coluna numérica (ano):



- Média e Mediana: Indicadores da tendência central das emissões de CO<sub>2</sub> por ano.
- Desvio Padrão e Variância: Medidas de dispersão que apontam para a variabilidade das emissões entre os diferentes países.
- Valores Mínimos e Máximos: Valores extremos que indicam os países com menores e maiores emissões em cada ano.
- Histograma de Distribuição: Visualização que permite detectar a existência de outliers e observar a distribuição dos dados em cada ano.

A análise estatística confirmou a presença de alguns países com emissões muito superiores aos demais, o que pode influenciar na interpretação dos resultados. Além disso, padrões de crescimento ou diminuição ao longo dos anos podem ser observados nas medidas de tendência central.

## **5. MODELO BASE**

Neste capítulo, será apresentado o desenvolvimento de um modelo de regressão linear aplicado às emissões totais de CO<sub>2</sub> ao longo dos anos, com o objetivo de criar uma análise de tendência para esse indicador ambiental. A construção deste modelo foi realizada em etapas que englobam a preparação dos dados, ajuste do modelo e visualização dos resultados.

```

1 total_emissions = data_teste.loc['Total'] # Seleciona a linha 'Total'
2
3 # Filtrando os valores de emissões apenas para os anos (colunas de 1990 a 2020, removendo 'Total Emissions')
4 total_emissions_anos = total_emissions.drop('Total Emissions')
5
6 # Convertendo os anos (que são as colunas) para inteiros, se necessário
7 anos = total_emissions_anos.index.astype(int)
8
9 # Criando o gráfico de linhas
10 plt.figure(figsize=(10, 6))
11 plt.plot(anos, total_emissions_anos.values, marker='o', label='Emissões Totais')
12
13 X = np.array(anos).reshape(-1, 1) # Convertendo anos para uma matriz 2D
14 X = sm.add_constant(X) # Adicionando intercepto
15 y = total_emissions_anos.values
16
17 model = sm.OLS(y, X).fit()
18
19 # Fazendo a previsão para todos os anos
20 predictions = model.predict(X)
21
22 # Plotando a linha de regressão
23 plt.plot(anos, predictions, color='red', label='Linha de Regressão')
24 # Configurando os rótulos dos eixos
25 plt.xlabel('Ano')
26 plt.ylabel('Emissões Totais')
27 plt.title('Evolução das Emissões Totais por Ano')
28
29 # Exibindo a legenda e o grid
30 plt.legend()
31 plt.grid(True)
32
33 # Mostrando o gráfico
34 plt.show()

```

*Figura 6 - O Modelo Base*

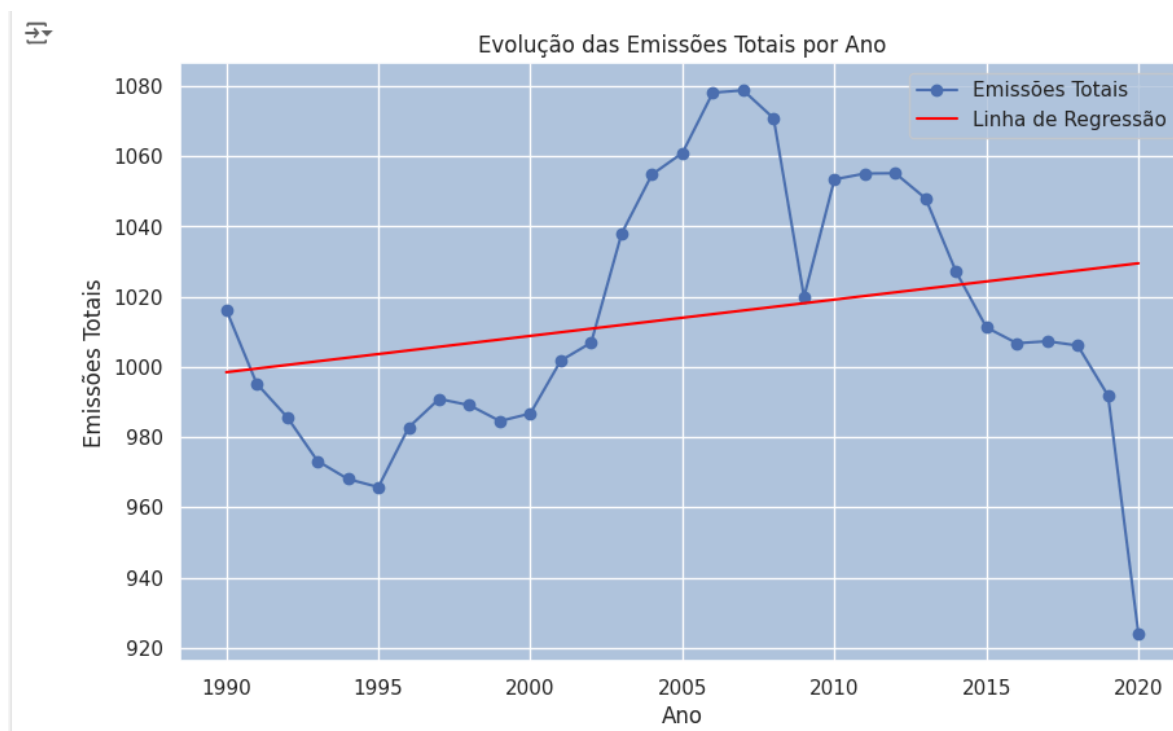
Inicialmente, os dados de emissões totais de CO<sub>2</sub> foram extraídos da linha chamada "Total" no conjunto de dados `data_teste`. Essa linha contém os valores das emissões acumuladas para cada ano disponível. Em seguida, os dados foram filtrados para incluir apenas os anos relevantes (de 1990 a 2020), removendo-se quaisquer colunas extras, como aquelas com o total acumulado geral. Para garantir que esses valores fossem interpretados corretamente pelo modelo de regressão, a indexação dos anos foi convertida para o tipo inteiro.

Após a preparação dos dados, o modelo de regressão linear foi estruturado. Primeiramente, os anos foram organizados em uma matriz bidimensional (necessária para compatibilidade com o modelo). A seguir, foi adicionada uma constante para o intercepto utilizando a função `sm.add_constant(X)`, em que `X` representa os anos convertidos para uma matriz 2D. As emissões totais de CO<sub>2</sub> de cada ano foram definidas como o vetor `y`, extraído previamente. Com essas definições, o modelo foi ajustado por meio do método

de Mínimos Quadrados Ordinários (OLS), através do comando `sm.OLS(y, X).fit()`. Esse ajuste permitiu que uma linha de regressão fosse determinada, representando a relação entre o ano e as emissões totais de CO<sub>2</sub> ao longo do tempo.

Após o treinamento, o modelo foi utilizado para prever as emissões totais para cada ano presente no conjunto de dados. Essas previsões foram geradas através do comando `model.predict(X)`, resultando em uma série de valores que representam a linha de tendência das emissões ao longo dos anos.

Com as previsões obtidas, foi construído um gráfico para visualização dos resultados. O gráfico incluiu as emissões totais observadas por ano, com pontos marcados, além da linha de regressão em vermelho, representando as previsões do modelo. Foram atribuídos rótulos aos eixos (Ano no eixo x e Emissões Totais no eixo y) e um título ao gráfico, intitulado "Evolução das Emissões Totais por Ano". A legenda e uma grade de fundo foram acrescentadas para melhorar a clareza do gráfico. Finalmente, o gráfico foi exibido com o comando `plt.show()`, possibilitando a visualização da tendência de crescimento nas emissões ao longo dos anos e da linha de regressão que estima essa tendência.



*Figura 7 - Gráfico de 'Evolução das emissões totais por Ano'*

Assim, foi realizada uma análise de tendência nas emissões totais de CO<sub>2</sub> ao longo do tempo, utilizando-se um modelo de regressão linear que permitiu a construção de uma linha de ajuste ideal. Essa linha de regressão fornece uma visão simplificada da direção geral dos dados, evidenciando um aumento nas emissões ao longo do período analisado. Desse modo, o modelo base apresentou uma análise inicial do comportamento temporal das emissões, servindo como ponto de partida para futuras melhorias e refinamentos.

O gráfico sugere que as emissões totais têm aumentado de forma consistente ao longo do período analisado, o que é preocupante do ponto de vista ambiental, uma vez que as emissões de gases de efeito estufa são um dos principais causadores das mudanças climáticas. No entanto, a análise também revela a existência de flutuações anuais e um possível ponto de inflexão, o que indica que a relação entre as emissões e o tempo pode ser mais complexa do que sugere o modelo de regressão linear.

## 2.1 CRONOGRAMA

Neste capítulo, apresenta-se o cronograma do projeto, organizado de forma a contemplar as principais etapas e prazos para a realização das tarefas, além de checkpoints e entregáveis fundamentais. O planejamento foi estruturado para permitir a flexibilização e monitoramento contínuo do progresso, com a previsão de reuniões semanais ou quinzenais que possibilitam um acompanhamento próximo do desenvolvimento. A natureza iterativa do processo de modelagem em machine learning também foi considerada, permitindo revisões e ajustes ao longo das etapas. Destaca-se, ainda, a importância do detalhamento da etapa 3, em que será especificada a escolha de modelos e métricas, garantindo que todos estejam alinhados quanto aos objetivos.

Etapa	Data de Início	Data de Término	Checkpoint	Entregáveis	Observações
Definição do Projeto e equipe	06/09	30/09	22/08	Conforme descrito	Reunião semanal para acompanhar o progresso.
Referencial teórico e cronograma	01/10	15/10	19/09	Conforme descrito	Iniciar a coleta e preparação dos dados.
Implementação parcial	16/10	28/10	17/10	Conforme descrito	Explorar diferentes modelos de machine learning.
Implementação e entrega final	29/10	18/11	-	Relatório final com resultados, código e apresentação.	Incluir análise de sensibilidade e incertezas.

### Flexibilização

Para garantir a flexibilidade necessária no acompanhamento do projeto, foram incluídas reuniões regulares que permitem uma supervisão constante e ágil. Esse processo facilita a identificação precoce de problemas e a correção de possíveis desvios em relação aos objetivos estabelecidos.

### **Detalhamento da Etapa 3**

A etapa de implementação parcial conta com um detalhamento específico para a escolha de modelos e métricas, etapa fundamental para o alinhamento da equipe e para garantir que todos os envolvidos tenham uma visão clara e coesa dos objetivos e dos métodos que serão aplicados.

### **Iteratividade**

A iteratividade do desenvolvimento de modelos de machine learning é reforçada ao longo de todo o cronograma. Esse enfoque permite que ajustes e refinamentos sejam realizados com base nos resultados obtidos nas etapas anteriores, aprimorando o desempenho do modelo final e assegurando a qualidade do produto final a ser entregue.

O cronograma aqui estabelecido serve como guia para o desenvolvimento do projeto, permitindo um monitoramento eficaz e um direcionamento claro em cada fase do trabalho.

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

- (1) POYTING, Mark; PRAZERES, Leandro. Mudanças climáticas foram 'principal' fator para seca recorde na Amazônia, diz estudo: o que isso significa para o futuro da floresta?. [S. l.], 24 jan. 2024. Disponível em:  
<https://www.bbc.com/portuguese/articles/c88nr0940j8o>. Acesso em: 11 set. 2024.
- (2) CO2 emissions (metric tons per capita) - Latin America & Caribbean. [S. l.], 14 fev. 2023. Disponível em:  
<https://data.worldbank.org/indicator/EN.ATM.CO2E.PC?end=2020&locations=ZJ&start=1990&view=chart>. Acesso em: 11 set. 2024.
- (3) RENEWABLE energy consumption (% of total final energy consumption) - Latin America & Caribbean. [S. l.], 11 abr. 2023. Disponível em:  
<https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS?locations=ZJ>. Acesso em: 11 set. 2024.
- (4) ARAUJO, Arney Rayol Moura de; LOBATO, Tarcísio da Costa. Análise e previsão das emissões de CO2 provenientes dos Combustíveis do Setor de Transporte Rodoviário no Brasil 1990- 2020. [S. l.], 1 dez. 2020. Disponível em:  
<http://repositorioinstitucional.uea.edu.br/bitstream/riuea/5280/1/Análise%20e%20previsão%20das%20emissões%20de%20CO2%20provenientes%20dos%20Combustíveis%20do%20Setor%20de%20Transporte%20Rodoviário%20no%20Brasil%201990-2020.pdf>. Acesso em: 30 set. 2024.
- (5) SOARES, Francisco. Aplicação de Machine Learning para Previsão das Emissões de CO2e Produzidas pela Produção de Energia no Estado do Amazonas: Um Estudo de Caso Usando R. [S. l.], 12 fev. 2024. Disponível em:  
<https://medium.com/@fanciscosoares/aplicação-de-machine-learning-para-previsão-das-emissões-de-co2e-produzidas-pela-produção-de-1817cdd34729>. Acesso em: 30 set. 2024.

## 5. APÊNDICES

(1) Link do repositório do projeto:

<https://github.com/ViniSegatto/ProjetoAplicadoIV/tree/main/Projeto>

(2) Código no Google Colaboratory:

 Untitled48.ipynb