

MLPRegressor (Aplicação em dados Climaticos)

Ana Paula Vanderley

- Multi-layer Perceptron Regressor é um modelo de regressão baseado em redes neurais artificiais uma poderosa ferramenta para modelagem de regressão para aprender relações complexas entre variáveis de entrada e saída, bastante flexível para lidar com uma variedade de problemas de regressão.
- Além de apenas fazer previsões, o MLPRegressor nos permite entender como cada variável de entrada contribui para as previsões finais. Isso é útil para insights meteorológicos e tomada de decisão baseada em dados.
- O objetivo desse trabalho foi demonstrar a utilização do MLPRegressor com as features selecionadas através de um modelo de Randon Florest processo que foi bastante útil porque permitiu capturar a complexidade das relações entre variáveis meteorológicas, proporcionando uma modelagem robusta e previsões mais precisas das rajadas máximas de vento com esse banco de dados de clima de Salvador.

BANCO DE DADOS

Out[]:

	Data Medicao	PRECIPITACAO TOTAL, DIARIO (AUT) (mm)	PRESSAO ATMOSFERICA MEDIA DIARIA (AUT) (mB)	TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C)	TEMPERATURA MAXIMA, DIARIA (AUT) (°C)	TEMPERATURA MEDIA, DIARIA (AUT) (°C)	TEI D
0	01/01/2022	0.0	1.006.533.333	218.625	30.4	26.445.833	
1	02/01/2022	0.0	10.071.875	21.675	30.8	26.729.167	
2	03/01/2022	0.0	1007.2	21.258.333	30.9	26.708.333	
3	04/01/2022	0.0	1006.1	21.733.333	28.9	250.625	
4	05/01/2022	8.8	10.055.375	21.970.833	31.2	26.7	

CONTAGEM DO BANCO DE DADOS

Out[]:

Data Medicao365

PRECIPITACAO TOTAL, DIARIO (AUT)(mm)365

PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)(mB)365

TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C)365

TEMPERATURA MAXIMA, DIARIA (AUT)(°C)365

TEMPERATURA MEDIA, DIARIA (AUT)(°C)365

TEMPERATURA MINIMA, DIARIA (AUT)(°C)365

UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%)365

UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%)365

VENTO, RAJADA MAXIMA DIARIA (AUT)(m/s)365

VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s)365

dtype: int64

MEDIDAS DESCRITIVAS

Out[]:

	PRECIPITACAO TOTAL, DIARIO (AUT)(mm)	TEMPERATURA MAXIMA, DIARIA (AUT)(°C)	TEMPERATURA MINIMA, DIARIA (AUT)(°C)	UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT) (%)	VENTO, RAJADA MAXIMA DIARIA (AUT) (m/s)
count	365.000000	365.000000	365.000000	365.000000	365.000000
mean	5.247671	29.115616	23.134795	64.175342	7.341644
std	10.994366	1.890724	1.435603	7.176499	1.813008
min	0.000000	23.700000	19.100000	45.000000	4.000000
25%	0.000000	27.700000	22.100000	60.000000	6.100000
50%	0.800000	29.300000	23.200000	63.000000	7.000000
75%	5.800000	30.600000	24.200000	68.000000	8.300000
max	100.200000	33.200000	26.000000	90.000000	14.600000

TIPO DE DADOS

Out[]:

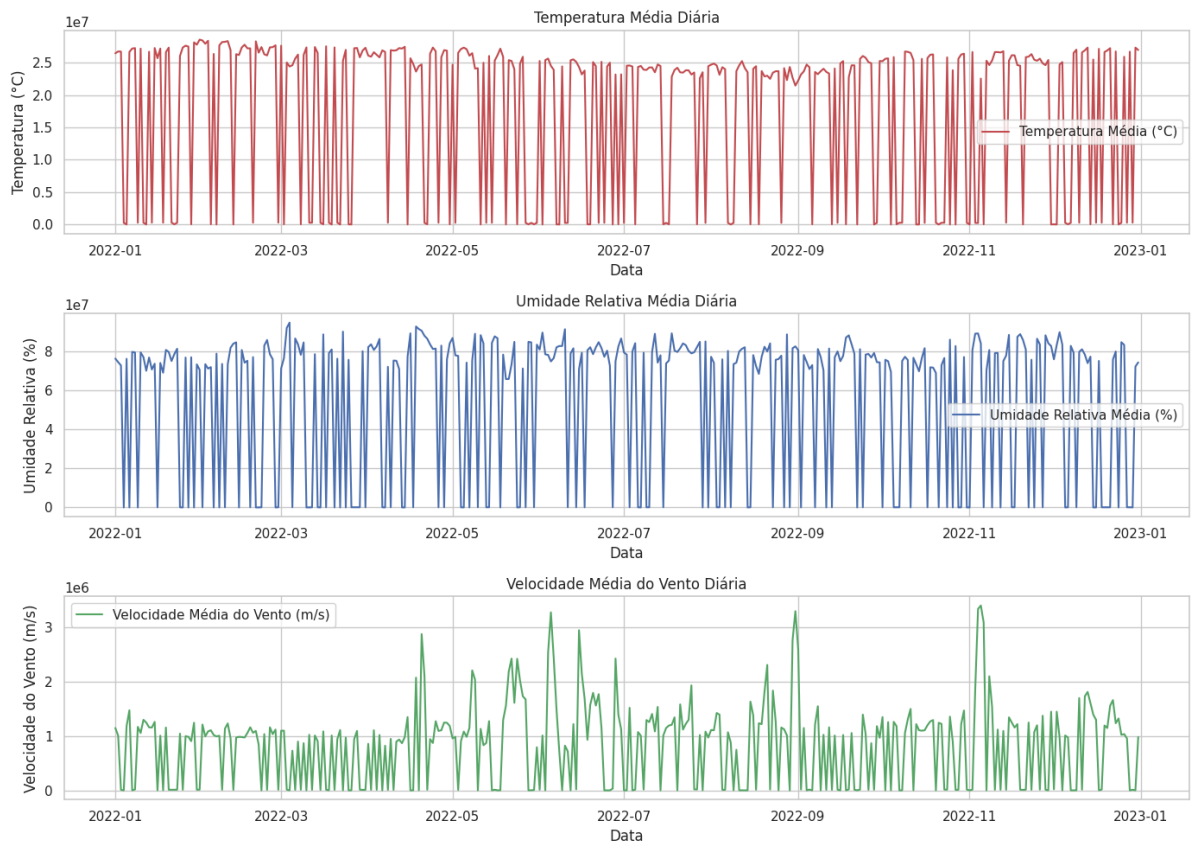
	Tipo
Data Medicao	datetime64[ns]
PRECIPITACAO TOTAL, DIARIO (AUT)(mm)	float64
PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)(mB)	object
TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C)	object
TEMPERATURA MAXIMA, DIARIA (AUT)(°C)	float64
TEMPERATURA MEDIA, DIARIA (AUT)(°C)	object
TEMPERATURA MINIMA, DIARIA (AUT)(°C)	float64
UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%)	object
UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%)	int64
VENTO, RAJADA MAXIMA DIARIA (AUT)(m/s)	float64
VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s)	object

- Foi preciso converter o tipo de dados de algumas variáveis para seguir com as análises .

Out[]:

	Tipo
Data Medicao	datetime64[ns]
PRECIPITACAO TOTAL, DIARIO (AUT)(mm)	float64
PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)(mB)	object
TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C)	float64
TEMPERATURA MAXIMA, DIARIA (AUT)(°C)	float64
TEMPERATURA MEDIA, DIARIA (AUT)(°C)	float64
TEMPERATURA MINIMA, DIARIA (AUT)(°C)	float64
UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%)	float64
UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%)	int64
VENTO, RAJADA MAXIMA DIARIA (AUT)(m/s)	float64
VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s)	float64

VISUALIZAÇÃO



VERIFICANDO VALORES AUSENTES

```
Out[ ]: Data Medicao                                0
PRECIPITACAO TOTAL, DIARIO (AUT)(mm)              0
PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)(mB)         0
TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C) 0
TEMPERATURA MAXIMA, DIARIA (AUT)(°C)              0
TEMPERATURA MEDIA, DIARIA (AUT)(°C)               0
TEMPERATURA MINIMA, DIARIA (AUT)(°C)              0
UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%)      0
UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%)     0
VENTO, RAJADA MAXIMA DIARIA (AUT)(m/s)            0
VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s)         0
dtype: int64
```

```
Out[ ]: Data Medicao                                0
PRECIPITACAO TOTAL, DIARIO (AUT)(mm)              0
PRESSAO ATMOSFERICA MEDIA DIARIA (AUT)(mB)         0
TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (AUT)(°C) 0
TEMPERATURA MAXIMA, DIARIA (AUT)(°C)              0
TEMPERATURA MEDIA, DIARIA (AUT)(°C)               0
TEMPERATURA MINIMA, DIARIA (AUT)(°C)              0
UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%)      0
UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%)     0
VENTO, RAJADA MAXIMA DIARIA (AUT)(m/s)            0
VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s)         0
dtype: int64
```

RANDOM FLOREST (Para selecionar as features mais importantes para o modelo)

DIVISÃO BASE DE TREINO E TESTE

ESCALONAMENTO DAS VARIAVEIS

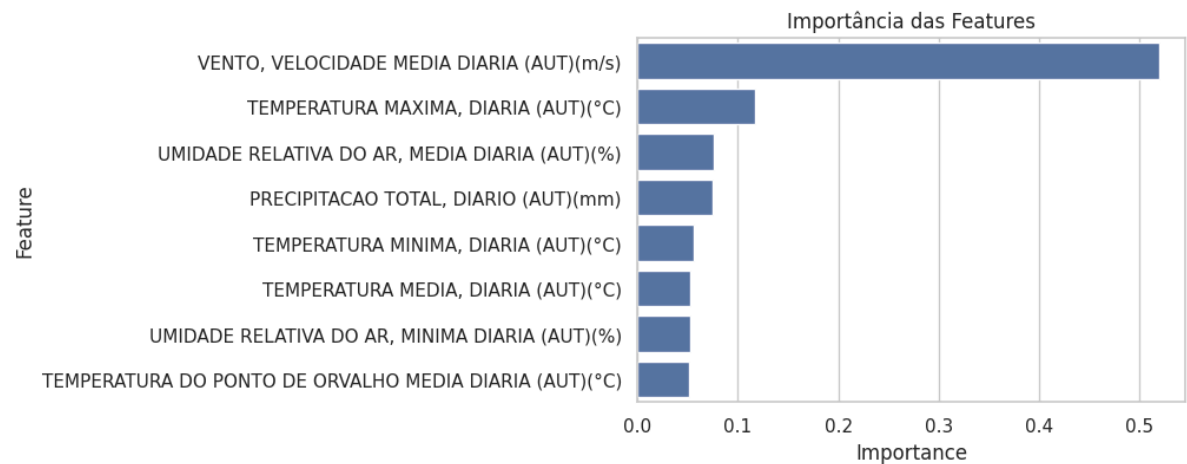
TREINANDO O MODELO

```
Out[ ]: ▼ RandomForestRegressor
RandomForestRegressor(random_state=30)
```

Criando um DataFrame com as importâncias das features

Out[]:

	Feature	Importance
7	VENTO, VELOCIDADE MEDIA DIARIA (AUT)(m/s)	0.519382
2	TEMPERATURA MAXIMA, DIARIA (AUT)(°C)	0.116908
5	UMIDADE RELATIVA DO AR, MEDIA DIARIA (AUT)(%)	0.076037
0	PRECIPITACAO TOTAL, DIARIO (AUT)(mm)	0.074873
4	TEMPERATURA MINIMA, DIARIA (AUT)(°C)	0.056202
3	TEMPERATURA MEDIA, DIARIA (AUT)(°C)	0.052729
6	UMIDADE RELATIVA DO AR, MINIMA DIARIA (AUT)(%)	0.052632
1	TEMPERATURA DO PONTO DE ORVALHO MEDIA DIARIA (...)	0.051238



SELECIONANDO AS FEACTURES MAIS INPORTANTES PARA O MODELO

- As features selecionadas foram VENTO, **VELOCIDADE MEDIA DIARIA (AUT)(m/s)** que representa a velocidade média do vento diário. É uma variável importante porque o comportamento do vento ao longo do dia pode influenciar diretamente a velocidade das rajadas máximas de vento e **TEMPERATURA MAXIMA, DIARIA (AUT)(°C)** que representa a temperatura máxima diária outra feature importante pois a temperatura pode afetar a dinâmica atmosférica e, consequentemente, influenciar a velocidade e intensidade do vento.

MODELO MLPRegressor

- O MLPRegressor é baseado em uma rede neural artificial com várias camadas de neurônios. Cada camada recebe entradas, realiza cálculos com pesos associados e passa esses valores para a próxima camada, até chegar à camada de saída que produz a previsão. As camadas intermediárias entre a entrada e a saída são chamadas de camadas ocultas.

- O modelo é treinado de forma iterativa usando métodos baseados em gradiente, Durante o treinamento, os pesos dos neurônios são ajustados para minimizar a função de perda, no caso da regressão é o erro quadrático médio (MSE).
- O MLPRegressor é adequado para capturar relações não lineares entre as variáveis de entrada (vento médio e temperatura máxima) e a variável de saída (rajada máxima de vento). Isso é importante porque os fenômenos atmosféricos, como o vento, frequentemente exibem comportamentos complexos e não lineares.
- O modelo também é capaz de aprender padrões complexos nos dados, o que é crucial quando se lida com previsões meteorológicas, onde múltiplas variáveis podem interagir de maneiras não óbvias para influenciar a velocidade das rajadas de vento.
- focamos aqui nesse trabalho em variáveis que têm um impacto direto ou indireto na variável de interesse (rajada máxima de vento). Isso nos permite que o modelo capture melhor as variações e padrões nos dados, resultando em previsões mais precisas.

ENCONTRANDO OS MELHORES PARAMETROS PARA O MODELO

- Aplicação do Grid Search para encontrar os melhores hiperparâmetros para o modelo de regressão usando rede neural MLP (Multi-layer Perceptron) ele define uma grade de possíveis combinações de hiperparâmetros para serem testadas.
- Foi utilizado o GridSearchCV para explorar todas as combinações possíveis de hiperparâmetros definidos em **param_grid**. O modelo é validado usando validação cruzada (cv=5) e a métrica de avaliação é o erro quadrático médio negativo (scoring='neg_mean_squared_error').

Melhores parâmetros encontrados:

```
{'alpha': 0.1, 'early_stopping': True, 'hidden_layer_sizes': (100, 80, 50, 30), 'max_iter': 500, 'n_iter_no_change': 30, 'validation_fraction': 0.1}
```

TREINAMENTO DO MODELO

Out[]:

```
MLPRegressor
MLPRegressor(alpha=0.1, early_stopping=True,
              hidden_layer_sizes=[100, 80, 50, 30], learning_rate='adaptive',
              max_iter=500, n_iter_no_change=30, random_state=30)
```

PREVISÕES

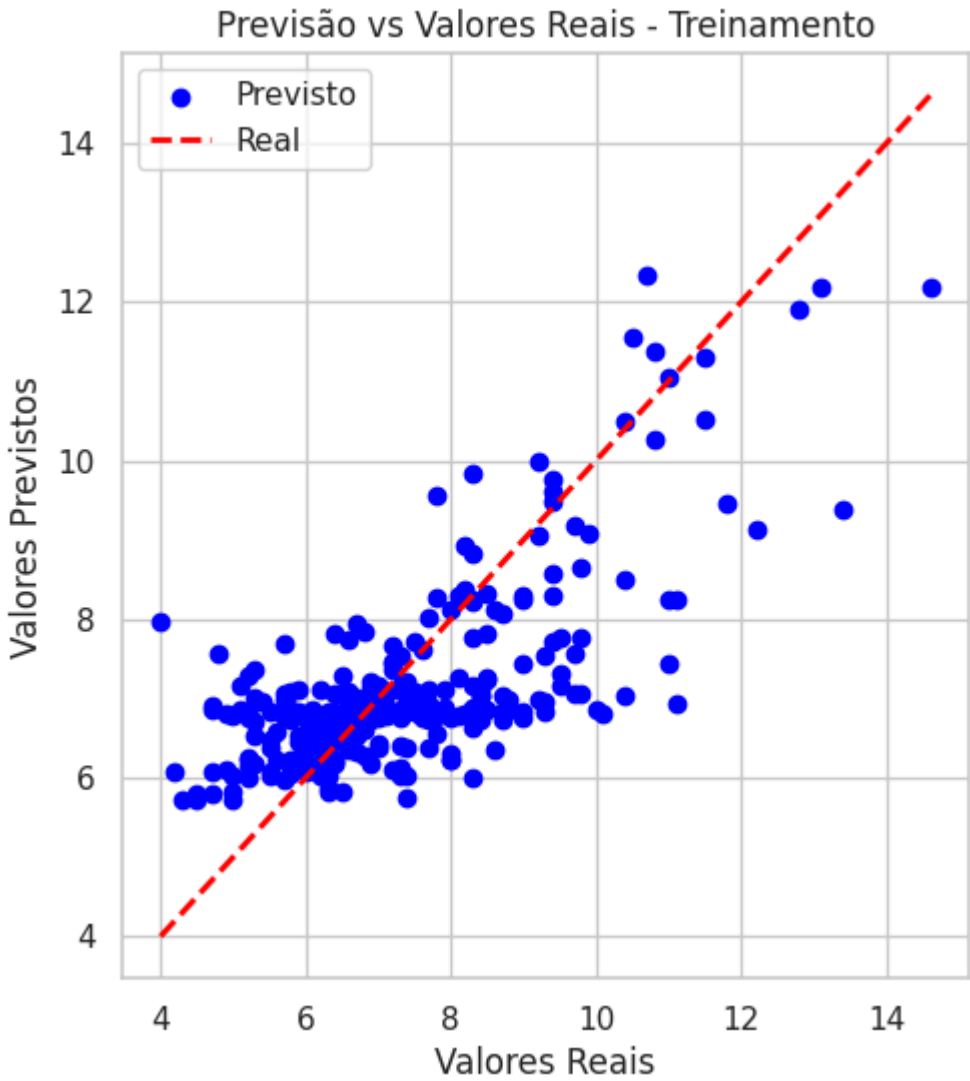
MEDIDAS DE DESEMPENHO DO MODELO

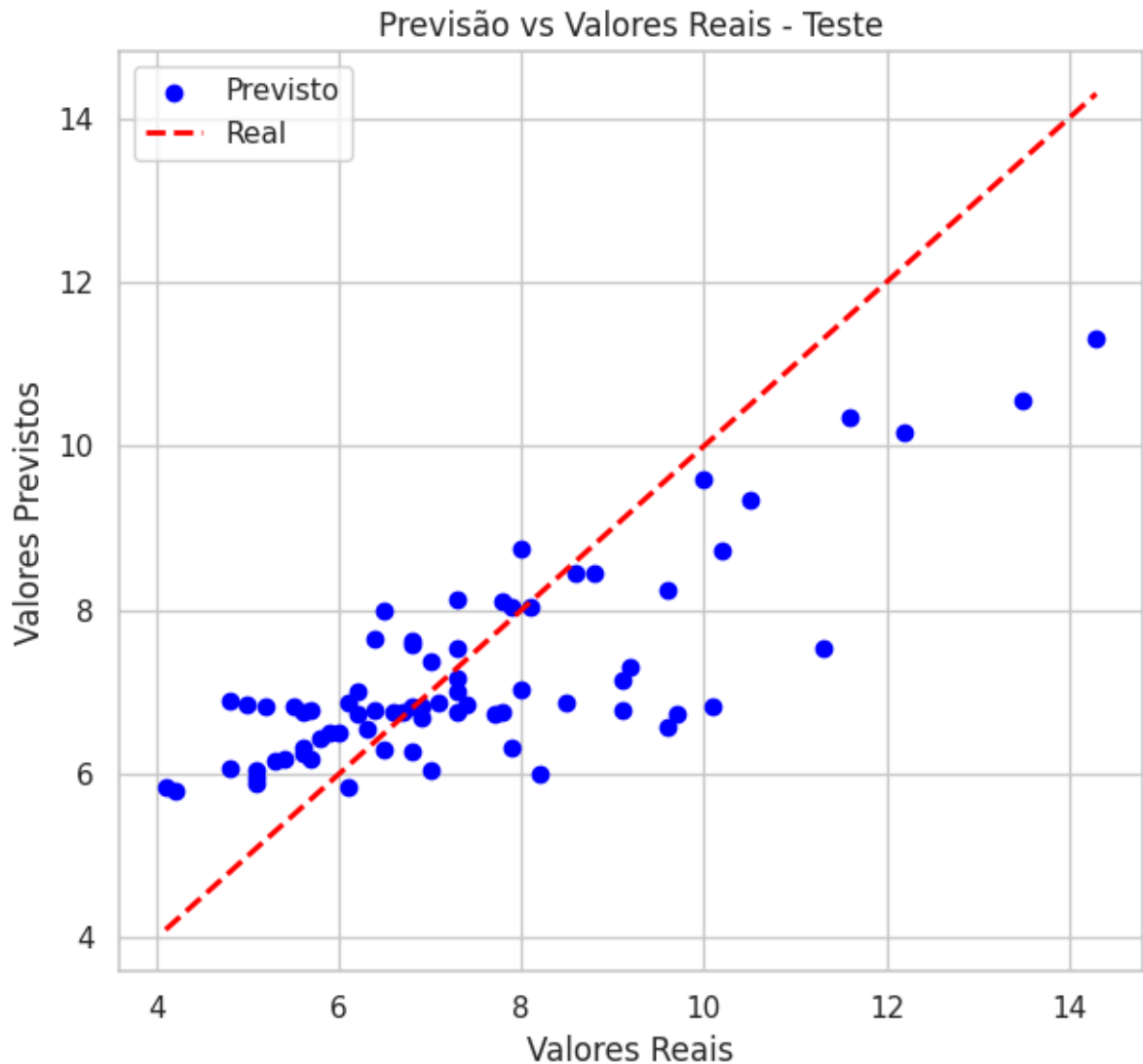
MSE Train: 2.0256, R^2 Train: 0.4906
MSE Test: 2.9781, R^2 Test: 0.5675

- Pela análise dos valores do MSE podemos ver quão próximas estão as previsões dos valores reais. No nosso caso tanto no conjunto de treino quanto no conjunto de teste, os valores de MSE estão relativamente próximos, nos indicando que o modelo está generalizando bem para novos dados.
- já o R^2 é uma medida que nos indica o quão bem as variáveis independentes explicam a variabilidade dos dados variando entre 0 a 1. No nosso caso tanto no treino quanto no teste, o R^2 está acima de 0.5, o que indica que o modelo consegue explicar uma parte significativa da variabilidade desses dados.
- Ou seja, o modelo pode ainda estar ligeiramente sobreajustado, mas ainda consegue generalizar razoavelmente bem para novos dados.
- lembrando que o modelo anteriormente foi treinado com 4 features e depois reduzido para 2 ,após a redução , houve uma melhoria notável nos resultados em comparação com os resultados anteriores, onde o R^2 tanto no treino quanto no teste aumentou, indicando um ajuste melhor do modelo aos dados.

PREVISTO X VALORES REAIS

Out[]: <matplotlib.legend.Legend at 0x7a27269630d0>





- Os resultados obtidos foram previsões mais precisas obtidas ao acrescentar variáveis significativas como a velocidade média do vento e a temperatura máxima, o modelo nos forneceu previsões mais precisas das rajadas máximas de vento, o que é crucial para aplicações como previsão meteorológica e gestão de riscos.
- Não somente para fazer previsões o MLPRegressor nos permitiu entender como cada variável de entrada contribui para as previsões finais. Fato esse bastante útil para insights meteorológicos e tomada de decisão baseada em dados pois foi possível capturar a complexidade das relações entre variáveis meteorológicas, proporcionando uma modelagem robusta e previsões mais precisas das rajadas máximas de vento em dados de clima em Salvador.