Regressão com Multi-Layer Perceptron

André Luiz Pires Guimarães

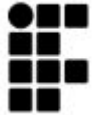
Daniel de Souza Miranda

Iago Nery Mendes

Lucas Elias de Andrade Cruvinel

Victor Hugo Brito da Silva Miranda

Avaliação Final - Deep Learning



Abordagem da apresentação

Apresentação inicial do problema
Pré-processamento realizado nos dados de produção de leite
Arquitetura utilizada e explicação da estrutura
Hiper Parametrização utilizando ”Grid-Search” e resultados
Estratégias utilizadas de “Data Augmentation” e resultados
Comparativos dos resultados utilizando abordagem anterior (RNN)

Apresentação inicial do problema

Predizer a produção de leite (*milk_production*) no dia a partir das variáveis dadas.

Principais pontos do problema:

1. Utilizar Regressão com Multi-layer Perceptron.
2. Treinar o modelo apenas com os 50 primeiros dias, e utilizar os restantes dias apenas para testes.

Dados utilizados

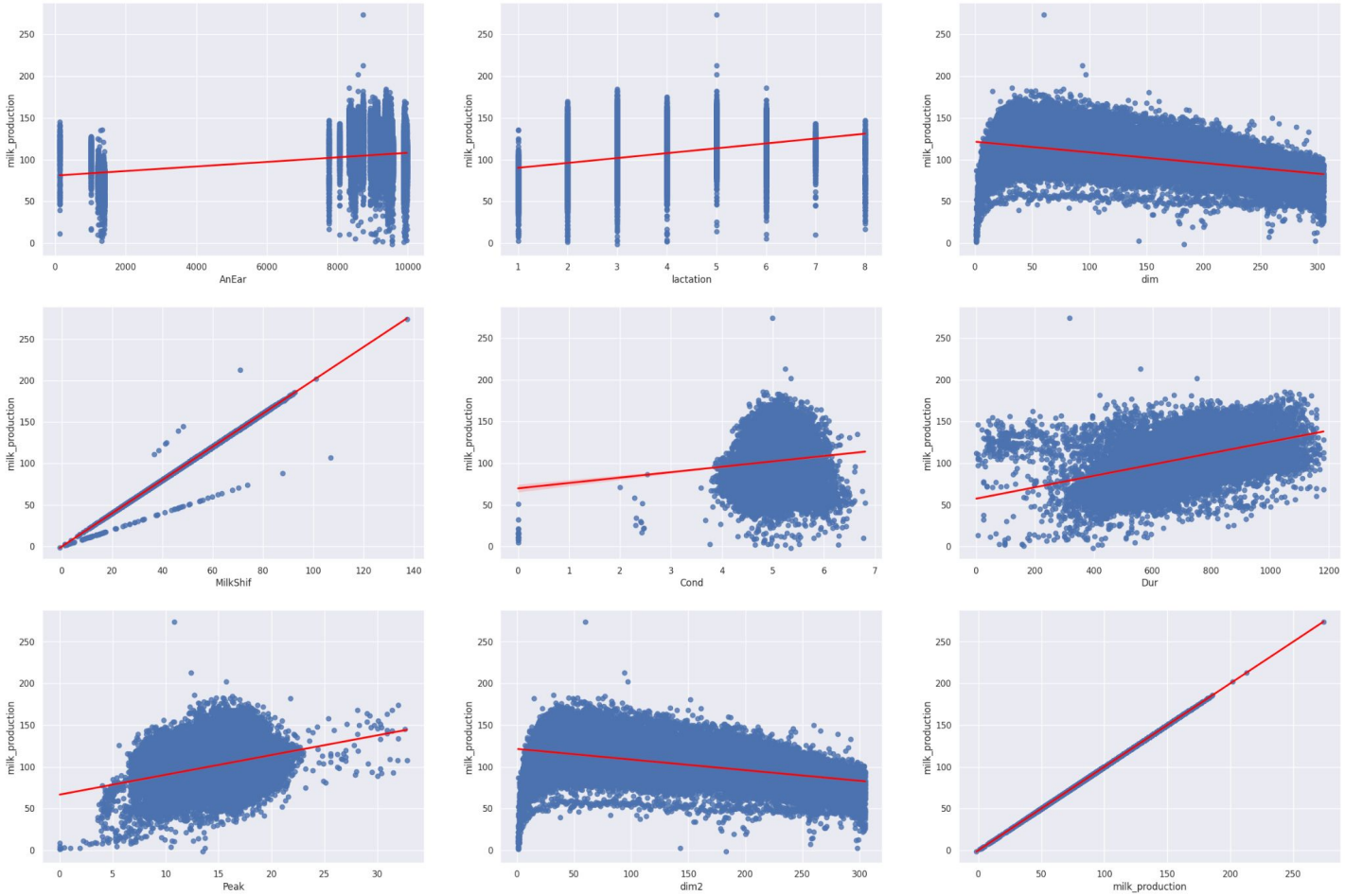
Os dados disponibilizados foram:

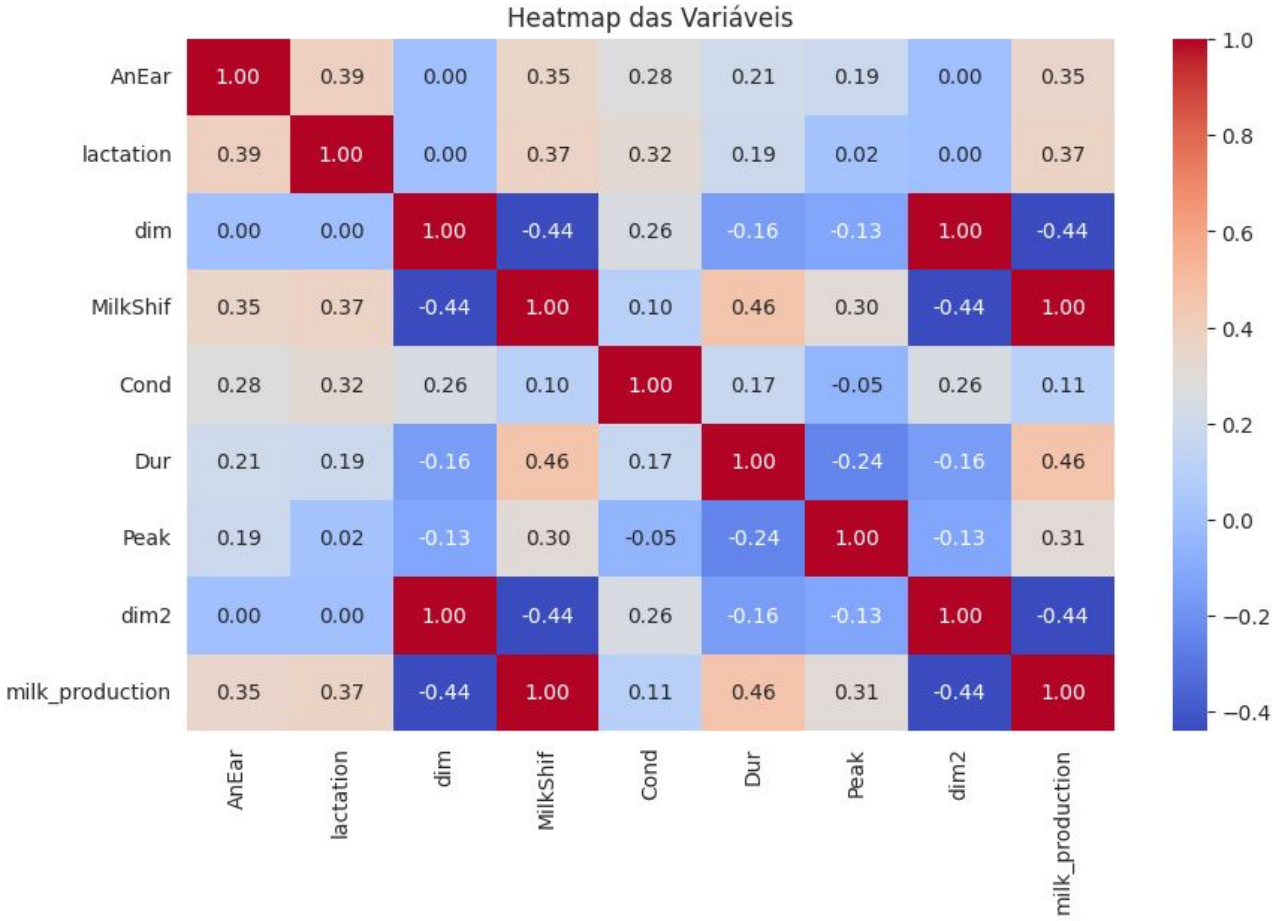
1. **AnEar**: código da vaca.
2. **Date**: data do registro da produção de leite.
3. **lactation**: quantos partos a vaca já teve.
4. **dim**: Dias após o último parto da vaca.
5. **milk_production**: produção de leite (em pounds - lbs).
6. **MilkShif**: Variação da produção.
7. **Cond**: condutividade elétrica no leite.
8. **Dur**: duração do aleitamento (em segundos).
9. **Peak**: Não especificado.
10. **Dim 2**: Não especificado.

Como **Peak** e **Dim 2** não foram especificados, não foram utilizados na modelagem.

Visualização dos dados

Relação das variáveis com o Produção de leite com uma reta de regressão




Heatmap das Variáveis
AnEar 1.00 0.39 0.00 0.35 0.28 0.21 0.19 0.00 0.35
lactation 0.39 1.00 0.00 0.37 0.32 0.19 0.02 0.00 0.37
dim 0.00 0.00 1.00 -0.44 0.26 -0.16 -0.13 1.00 -0.44
MilkShif 0.35 0.37 -0.44 1.00 0.10 0.46 0.30 -0.44 1.00
Cond 0.28 0.32 0.26 0.10 1.00 0.17 -0.05 0.26 0.11
Dur 0.21 0.19 -0.16 0.46 0.17 1.00 -0.24 -0.16 0.46
Peak 0.19 0.02 -0.13 0.30 -0.05 -0.24 1.00 -0.13 0.31
dim2 0.00 0.00 1.00 -0.44 0.26 -0.16 -0.13 1.00 -0.44
milk_production 0.35 0.37 -0.44 1.00 0.11 0.46 0.31 -0.44 1.00
AnEar
lactation
dim
MilkShif
Cond
Dur
Peak
dim2
milk_production
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Organização do dados

- Pré-processamento:
  - Remoção das features: **Date**, **Cond**, **Peak**, **Dim 2**.
  - Remoção dos registros com valores nulos.
  - Aplicação da normalização **MinMax Scale**.
- Divisão dos dados:
  - Os 50 primeiros dias para treino e validação, e os restantes para teste.
  - Desses 50 dias, 70% foram designados para treino e 30% para validação.
  - Logo, 35 dias para treino e 15 para validações.

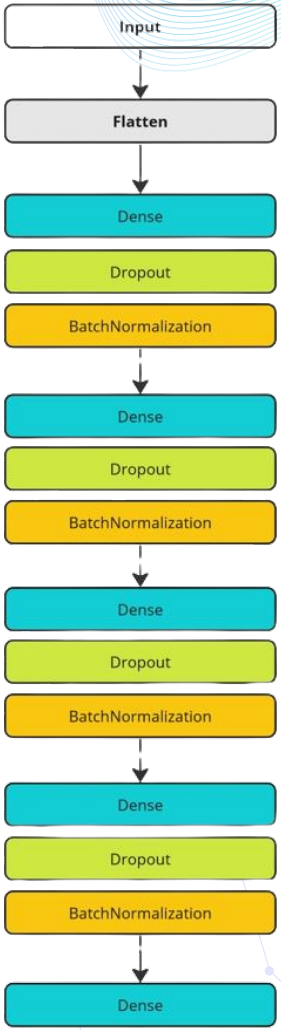
Arquitetura do modelo

O modelo foi arquitetado em formato sequencial, seguindo:

1. Camada com 4 entradas;
2. Camada Flatten para vetorização unidimensional da entrada.
3. O seguinte bloco que se repete 4 vezes:
    a. Camada densa com x neurônios;
    b. Camada de Dropout;
    c. Camada de Batch Normalization;
4. Camada de saída com um neurônio.

As camadas densas intermediárias são iniciadas com o inicializador “he_uniform”,
e função de ativação “ReLu”.

A camada final foi implementada com um função de ativação linear.



Arquitetura do modelo

Motivações de nossas escolhas:

- **ReLu**;
  - Prevenção do desaparecimento de gradiente.
- **He uniform**;
  - Otimizar desempenho e aumentar estabilidade da rede e dos gradientes.
- **Quatro blocos densos**;
  - Devido à complexidade do problema e reduzir Overfitting.
- **Dropout**;
  - Melhora a generalização e reduz Overfitting.
- **Batch normalization**;
  - Redução do problema de desaparecimento de gradiente e aumento da robustez.

Arquitetura do modelo

Parâmetros de treinamento

- **Otimizador:** Adam;
  - Otimizador estável, adaptativo e eficaz.

- **Métrica de Loss:** Erro médio Absoluto (MAE);
  - Insensível a outliers.
  - Penaliza erros grandes e pequenos

Ferramentas utilizadas

As ferramentas de *machine learning* utilizadas foram:

- Tensorflow;
- Keras;
- Scikit Learn;
- Scikeras.

Hiperparâmetros

Para o treinamento do modelo, foi necessário escolher os seguintes hiperparâmetros:

1. **Tamanho** das camadas ocultas:
   a. Quantidade de neurônios em cada camada densa intermediária.

2. Porcentagem de **Dropout**:
   a. Porcentagem de neurônios ignorados.

3. Tamanho de **batch**:
   a. Tamanho do batch por ciclo.

4. Taxa de **aprendizado**:
   a. Tamanho do passo que o modelo dá na direção correta durante o treinamento.

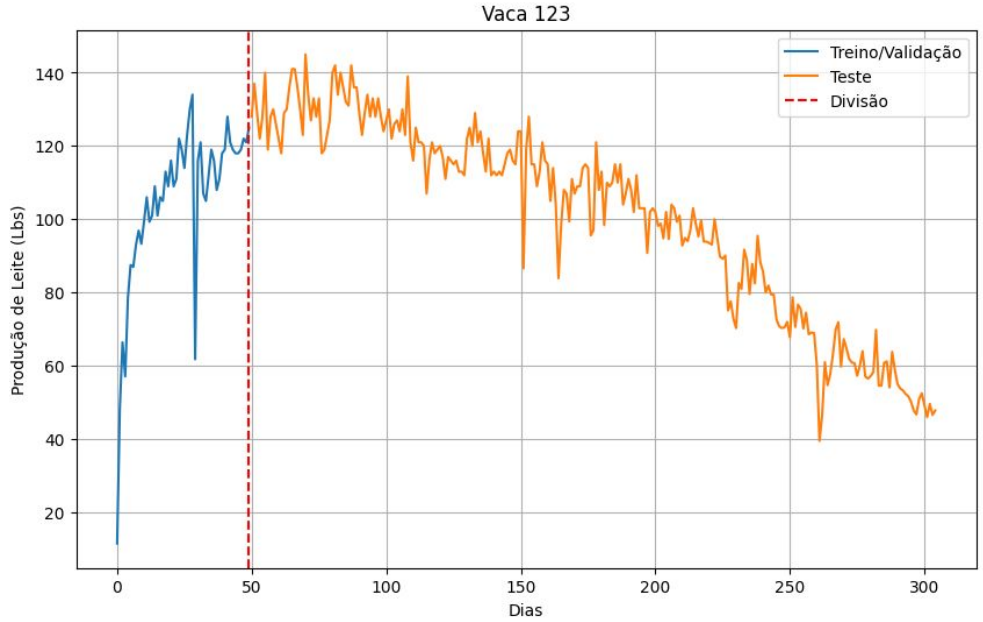
Hiperparâmetros

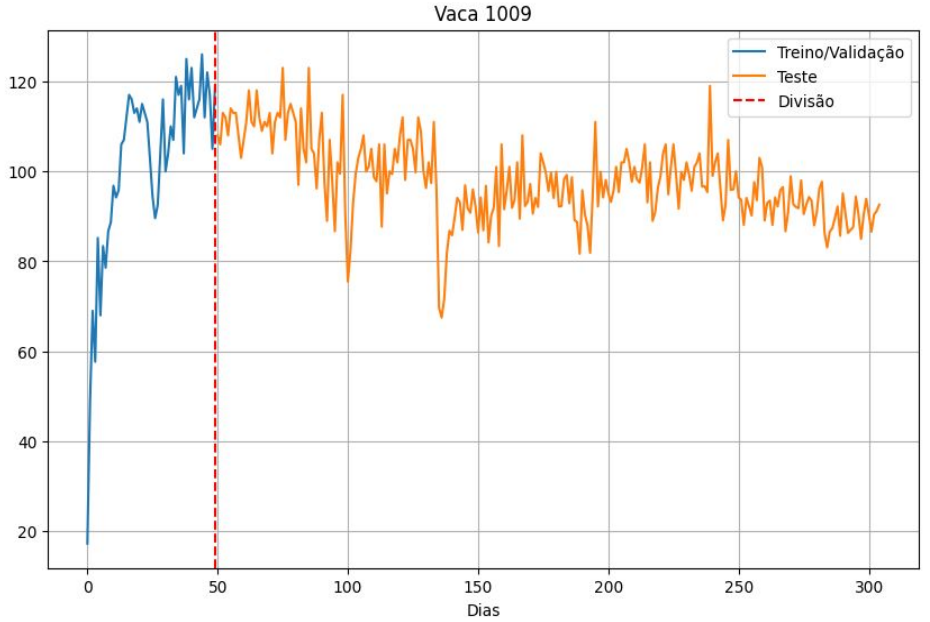
Para determinar quais valores de hiperparâmetros, foram inicialmente realizados testes avulsos, com valores aleatórios.
A partir disso, foram escolhidos certos limites de valores e utilizado o **Grid Search** para escolher os melhores valores.

1. **Tamanho** das camadas ocultas: 256, 512 ou 784 neurônios.
2. Porcentagem de **Dropout**: 30%, 40% ou 50%
3. Tamanho de **batch**: 16 ou 32 registros por batch.
4. Taxa de **aprendizado**: 0.01, 0.001 ou 0.0001

Totalizando 54 variações para treinamento.

Um grande problema





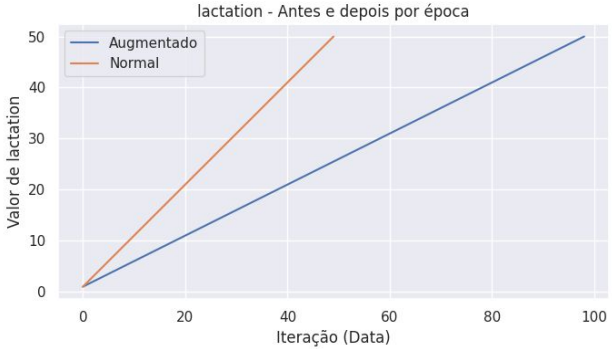
Incremento de dados com Interpolação

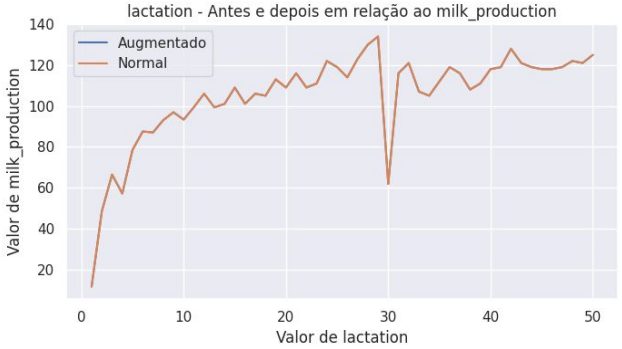
- Objetivo:
  - Verificar se incrementar a quantidade de dados de treino pode melhorar o resultado final.
- Método:
  - Técnica de interpolação linear.
- Algoritmo:
  - Organizar os dados em ordem temporal;
  - Introduzir uma nova linha de dados entre cada linha existente, calculando um ponto médio entre os valores.
- Exemplo com a vaca 123:
  - Antes:

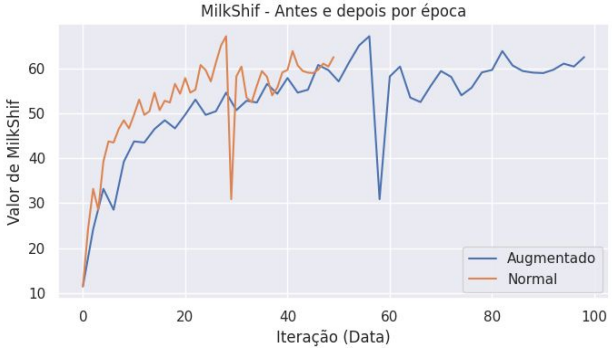
    registro[0] = [4. , 1. , 11.5 , 11.5 , 276.],
    registro[1] = [4. , 2. , 48.4 , 24.2 , 384.],
  - Depois:

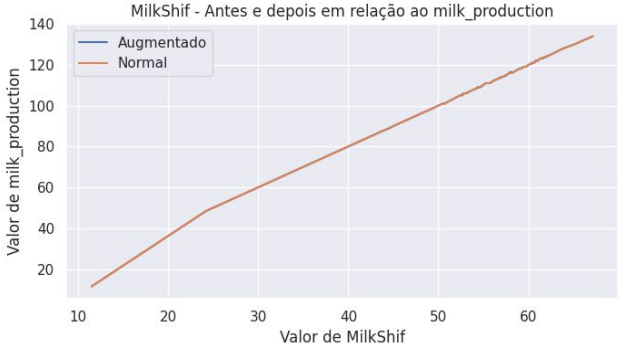
    registro[0] = [4. , 1. , 11.5 , 11.5 , 276.],
    registro[1] = [4. , 1.5 , 29.95 , 17.85 , 330.],
    registro[2] = [4. , 2. , 48.4 , 24.2 , 384.],

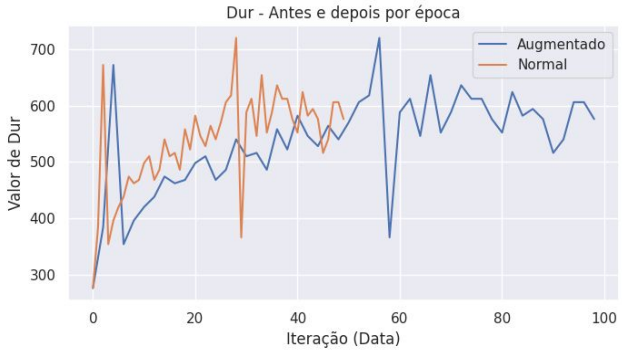
Visualização dos dados incrementados

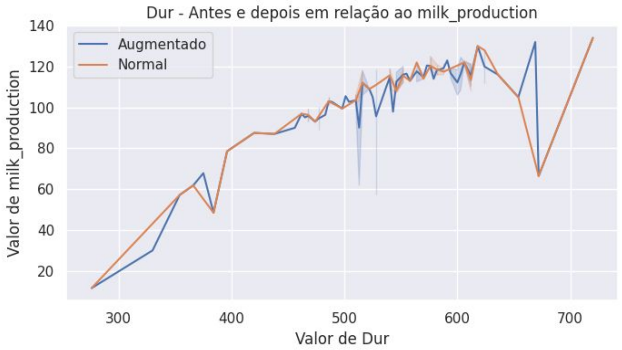

lactation - Antes e depois por época
Augmentado
Normal
Valor de lactation
Iteração (Data)


lactation - Antes e depois em relação ao milk_production
Augmentado
Normal
Valor de milk_production
Valor de lactation


MilkShif - Antes e depois por época
Augmentado
Normal
Valor de MilkShif
Iteração (Data)


MilkShif - Antes e depois em relação ao milk_production
Augmentado
Normal
Valor de milk_production
Valor de MilkShif


Dur - Antes e depois por época
Augmentado
Normal
Valor de Dur
Iteração (Data)


Dur - Antes e depois em relação ao milk_production
Augmentado
Normal
Valor de milk_production
Valor de Dur

Resultados

Hiperparâmetros

Os valores que o Grid Search indica como ideias são:

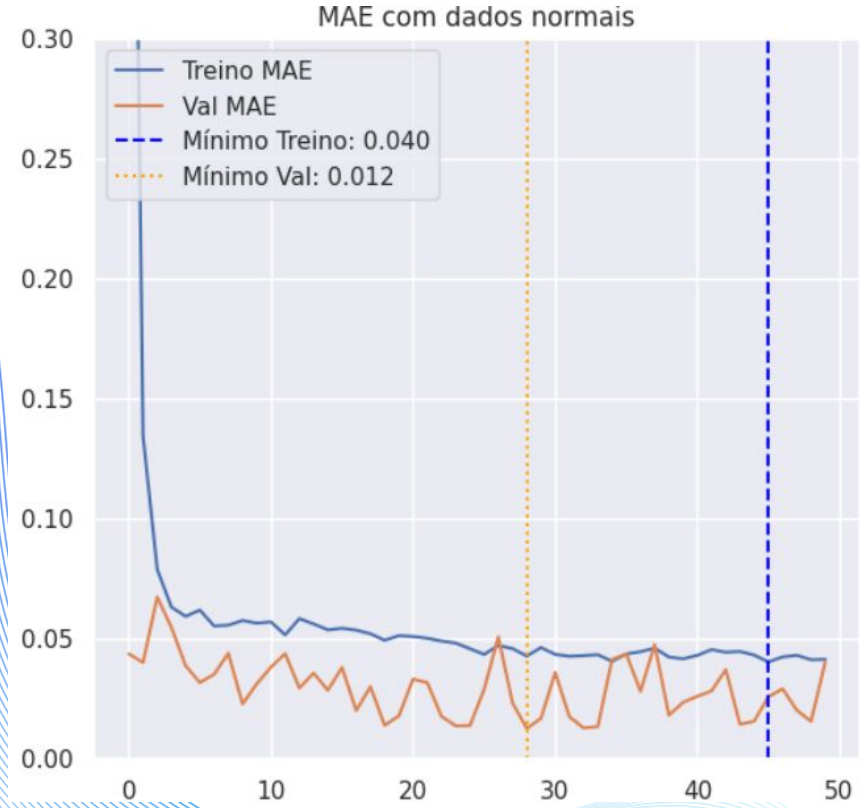
Com os dados normais:

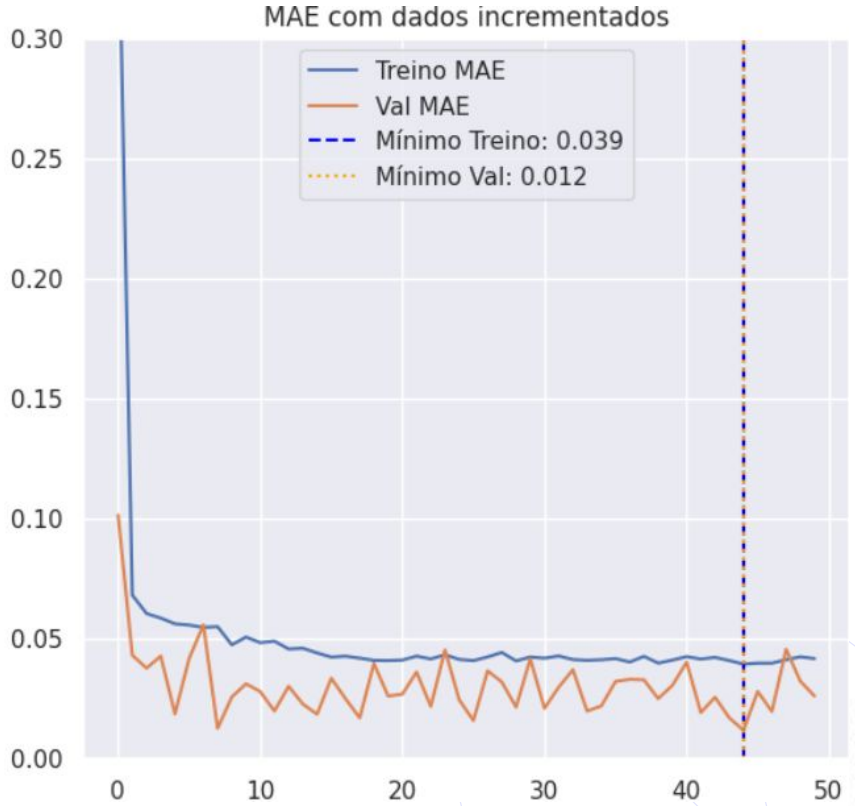
- **Tamanho** das camadas ocultas: 256 neurônios.
- Porcentagem de **Dropout**: 30%
- Tamanho de **batch**: 16 registros por batch.
- Taxa de **aprendizado**: 0.01

Com os data augmentation:

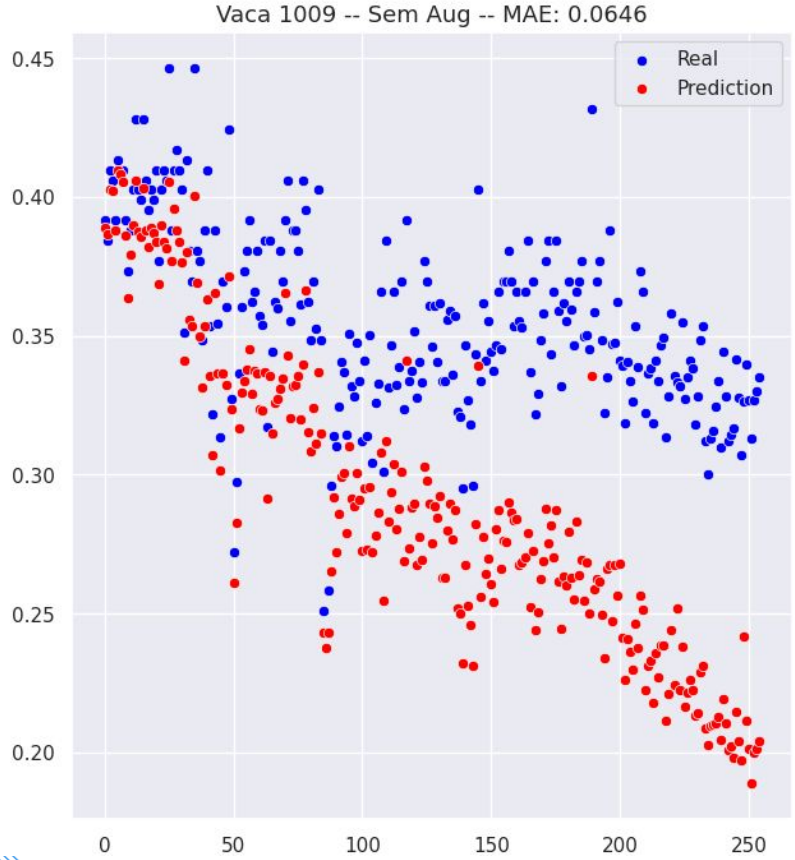
- **Tamanho** das camadas ocultas: 512 neurônios.
- Porcentagem de **Dropout**: 30%
- Tamanho de **batch**: 32 registros por batch.
- Taxa de **aprendizado**: 0.001

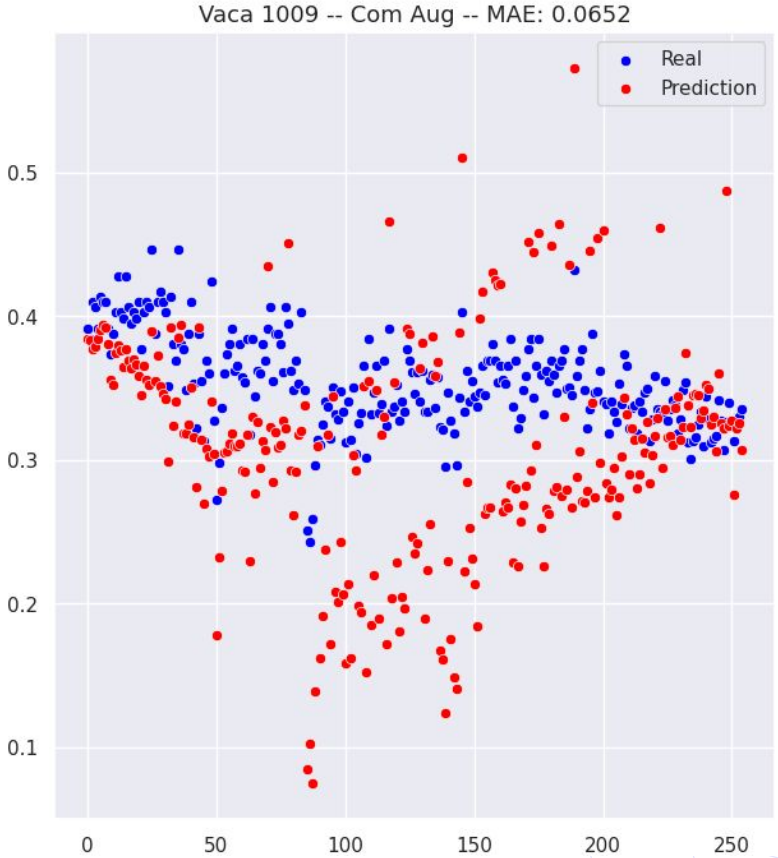
Treinamento e Validação

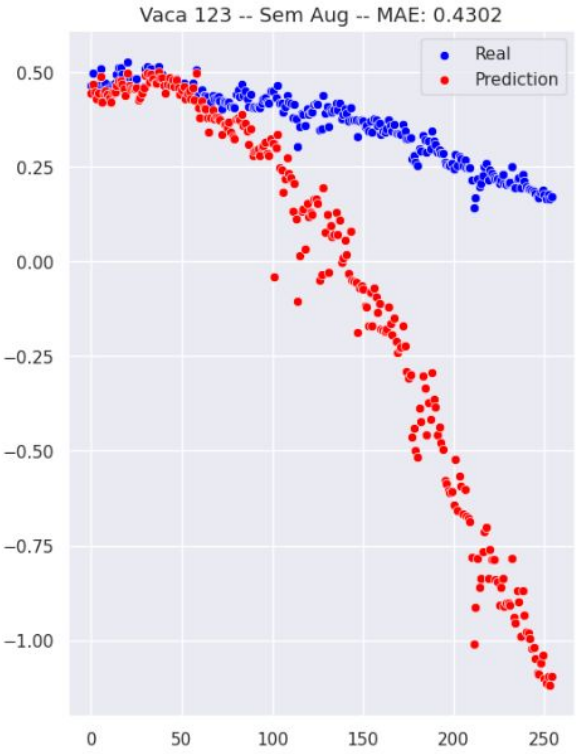


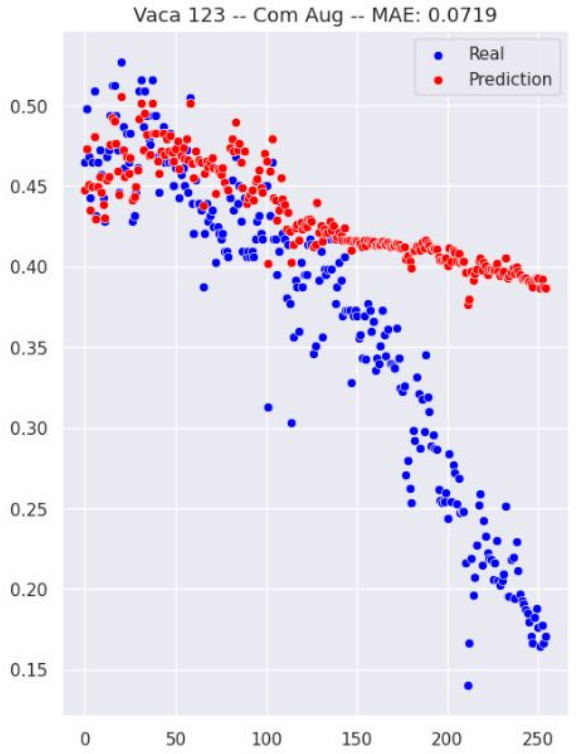


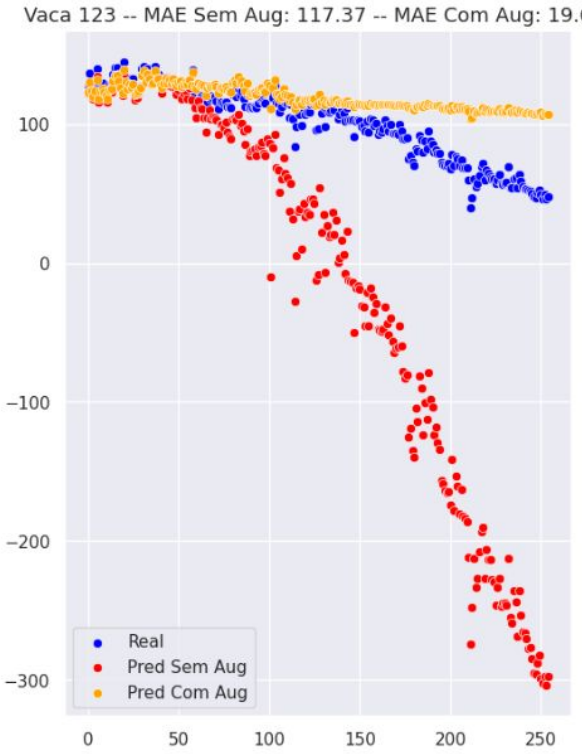
Comparando Vários Resultados

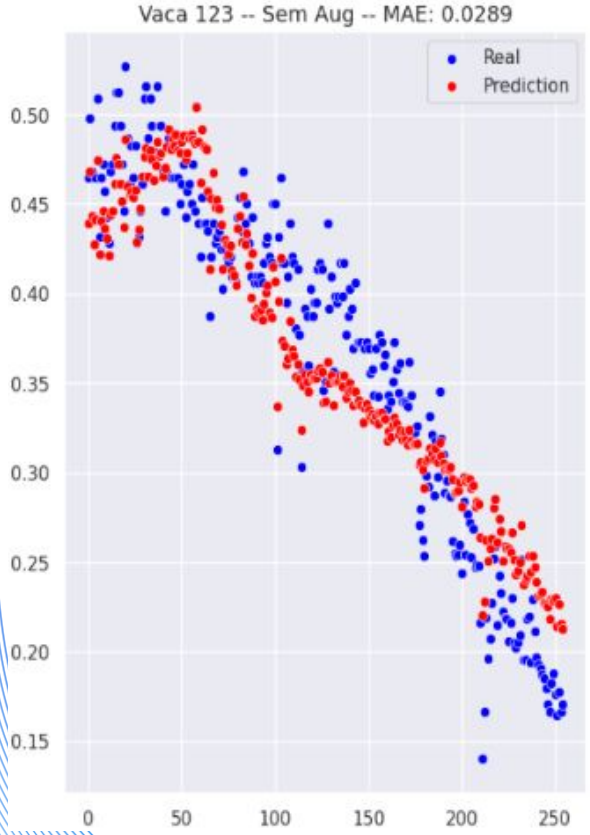

Vaca 1009 -- Sem Aug -- MAE: 0.0646
Real
Prediction
0.45
0.40
0.35
0.30
0.25
0.20
0
50
100
150
200
250

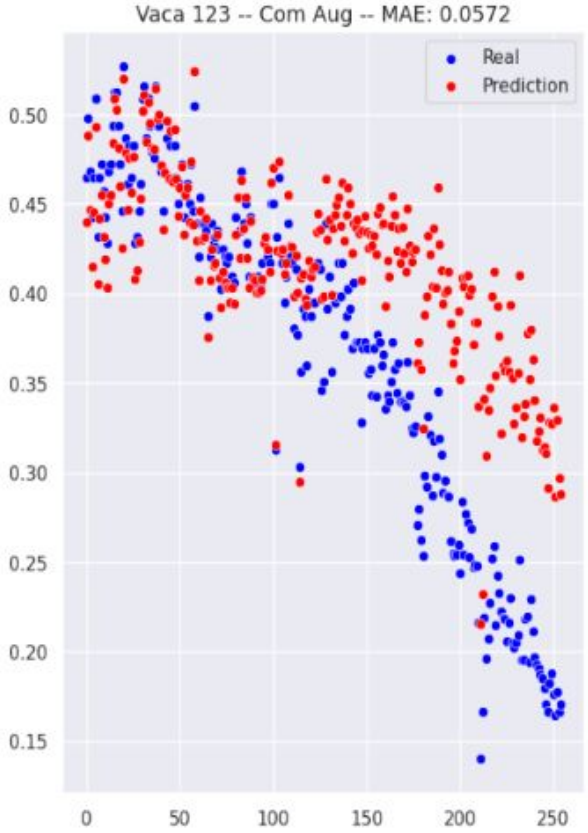

Vaca 1009 -- Com Aug -- MAE: 0.0652
Real
Prediction
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250

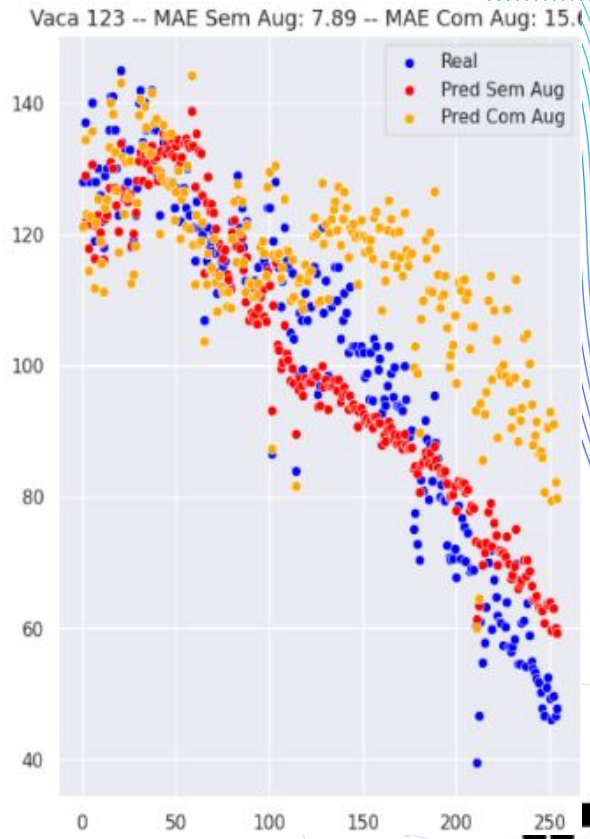

Vaca 123 -- Sem Aug -- MAE: 0.4302
Real
Prediction
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
50
100
150
200
250

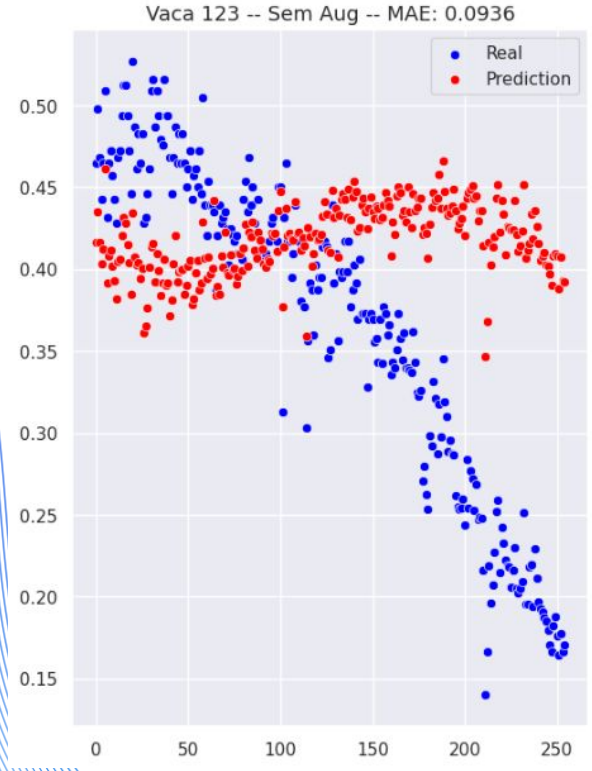

Vaca 123 -- Com Aug -- MAE: 0.0719
Real
Prediction
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
50
100
150
200
250

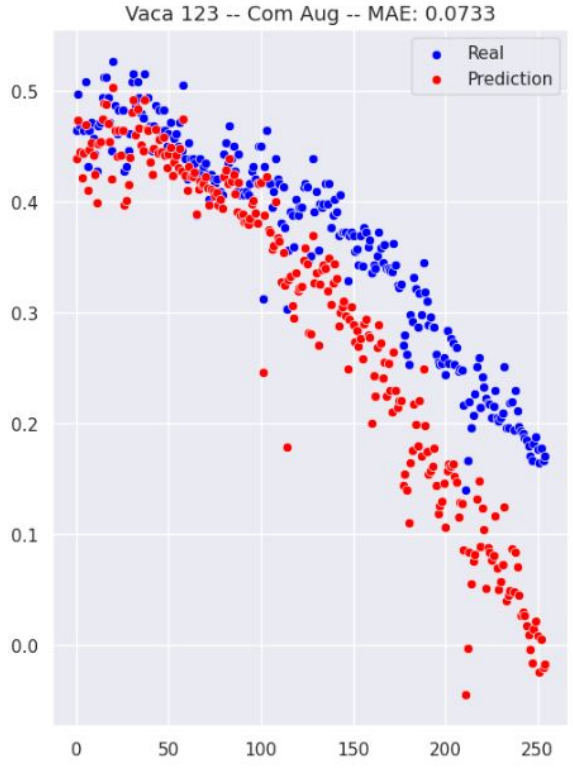

Vaca 123 -- MAE Sem Aug: 117.37 -- MAE Com Aug: 19.63
Real
Pred Sem Aug
Pred Com Aug
100
0
−100
−200
−300
0
50
100
150
200
250

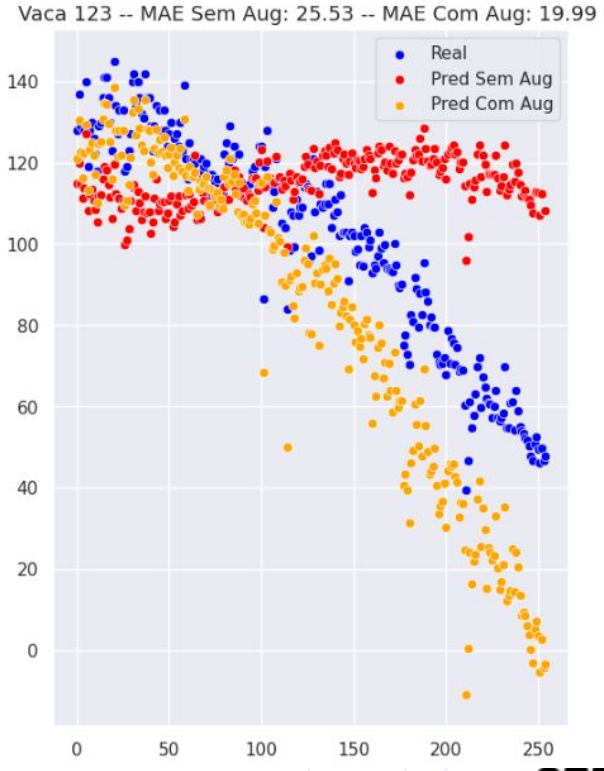

Vaca 123 -- Sem Aug -- MAE: 0.0289
Real
Prediction
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
50
100
150
200
250


Vaca 123 -- Com Aug -- MAE: 0.0572
Real
Prediction
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
50
100
150
200
250


Vaca 123 -- MAE Sem Aug: 7.89 -- MAE Com Aug: 15.
Real
Pred Sem Aug
Pred Com Aug
140
120
100
80
60
40
0
50
100
150
200
250


Vaca 123 -- Sem Aug -- MAE: 0.0936
Real
Prediction
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
50
100
150
200
250


Vaca 123 -- Com Aug -- MAE: 0.0733
Real
Prediction
0.5
0.4
0.3
0.2
0.1
0.0
0
50
100
150
200
250


Vaca 123 -- MAE Sem Aug: 25.53 -- MAE Com Aug: 19.99
Real
Pred Sem Aug
Pred Com Aug
140
120
100
80
60
40
20
0
0
50
100
150
200
250

Analise dos resultados

A análise dos resultados revelou uma **inconsistência** significativa, indicando que o modelo é altamente sensível à inicialização dos pesos.

Para mitigar essa instabilidade, foi implementada uma validação cruzada **K Fold** com 3 folds durante o Grid Search.

Essa técnica permite avaliar o desempenho do modelo em diferentes subconjuntos de dados, proporcionando uma estimativa mais robusta e confiável do desempenho geral.

A média dos resultados obtidos em cada fold oferece uma representação mais precisa do desempenho do modelo, reduzindo a influência da inicialização aleatória dos pesos.

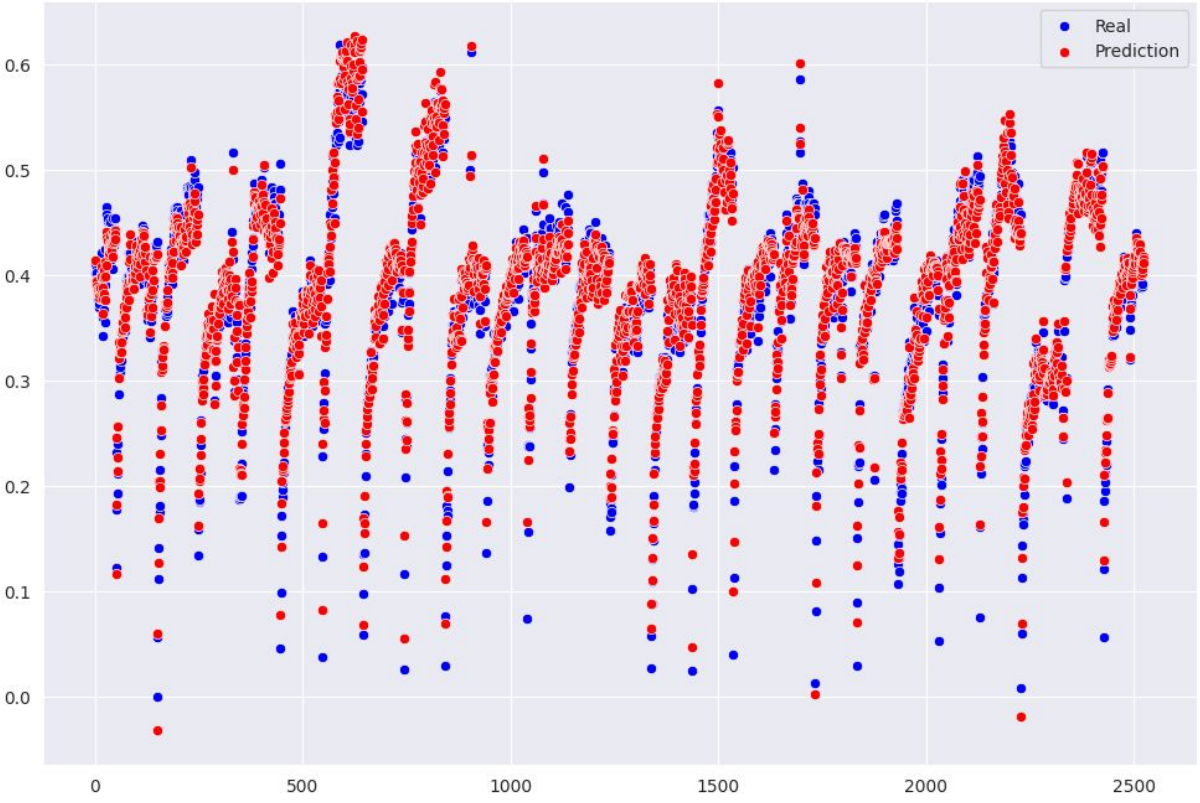
Em nossos experimentos, foi escolhido um tamanho de 3 folds.

Análise dos melhores resultados

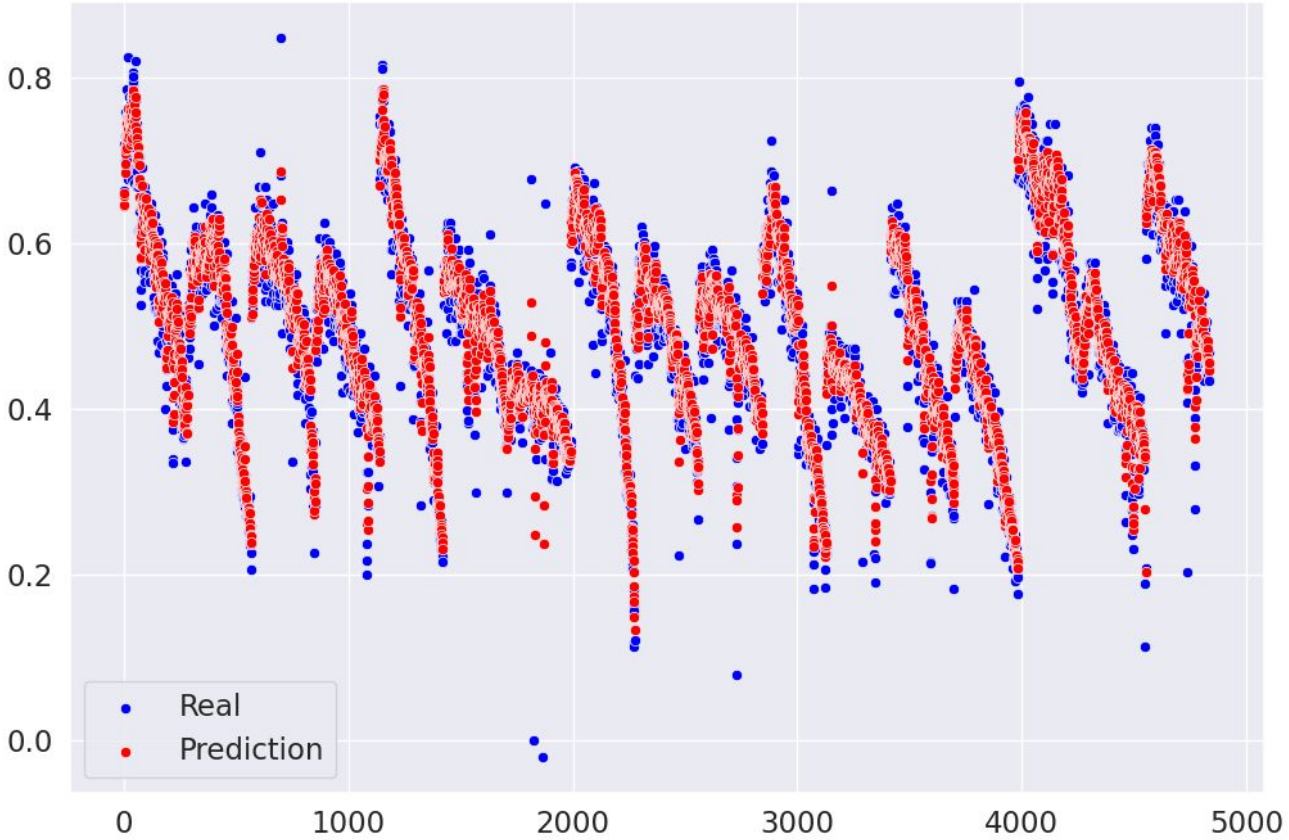
Augment	Quant. Neurônios	Learning Rate	Dropout	Batch Size	Média do KFold (MAE)
SEM	256	0.01	30%	16	0.0289
COM	512	0.001	30%	32	0.0341
COM	512	0.001	40%	32	0.0581
SEM	256	0.001	30%	16	0.0733
COM	256	0.001	40%	32	0.0795

Comparação com os resultados da atividade de RNN

MLP MAE: 0.0107 no conjunto de validação



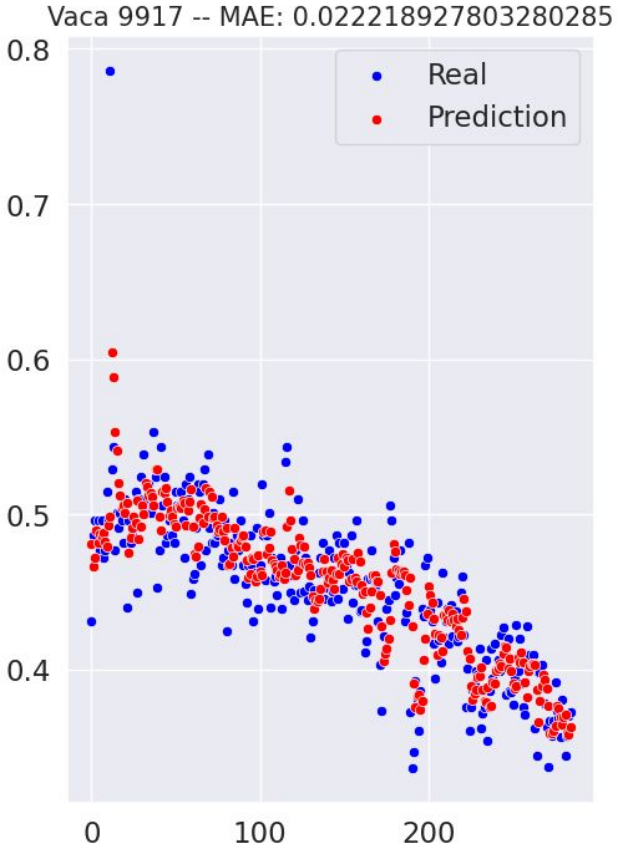
RNN MAE: 0.0251 no conjunto de validação

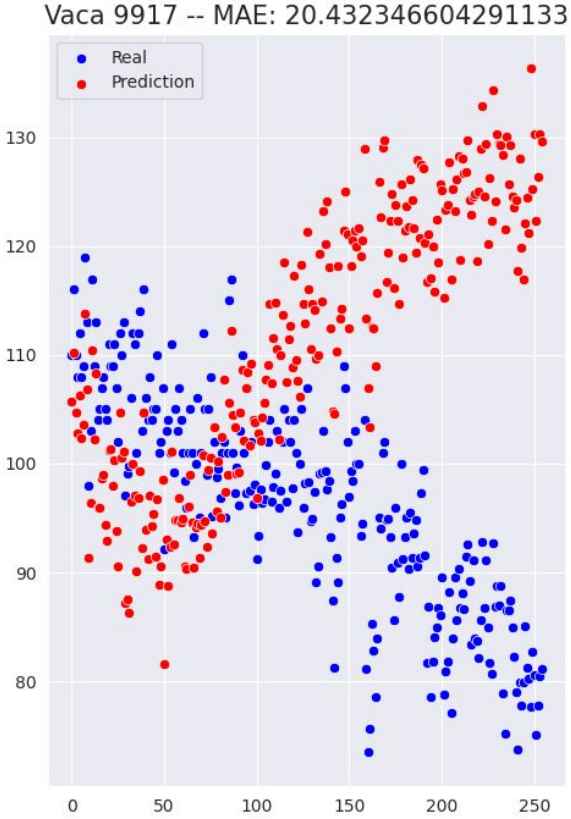


RNN vs MLP

RNN MAE: 0.0235 (no conjunto de teste)
MLP MAE: 0.0289 (sem Data Augmentation)
MLP MAE: 0.0341 (com Data Augmentation)

RNN vs MLP


Vaca 9917 -- MAE: 0.022218927803280285
Real
Prediction
0.8
0.7
0.6
0.5
0.4
0
100
200


Vaca 9917 -- MAE: 20.432346604291133
Real
Prediction
130
120
110
100
90
80
0
50
100
150
200
250

Conclusões

- O modelo implementado teve resultados melhores do que inicialmente previstos, mesmo sem o incremento de dados.
- O incremento de dados não teve impacto significativo, pois com ou sem o modelo chega a resultados similares.
- Através da hiper parametrização com Grid Search, foi possível chegar em um modelo que conseguisse resultados semelhantes a um modelo específico, como o modelo RNN da outra atividade.
- Embora tenha resultados ótimos, esse modelo é muito sensível à inicialização. Fazendo com que sua performance seja inconsistente de treinamento para treinamento.

Obrigado!