



## Encontro Técnico **AESABESP**

Congresso Nacional  
de Saneamento e  
Meio Ambiente

**34ETC-06304**

# APLICAÇÃO DE MODELOS PREDITIVOS DE MACHINE LEARNING PARA PARÂMETROS DE QUALIDADE EM ESTAÇÕES DE TRATAMENTO DE EFLUENTES

**Juliana Neves**

João Vítor Rios Fuck, Maria Alice Prado Cechinel, Ricardo Tristão, Rodrigo Campos de Andrade

**HydroInfo.** Endereço: Rua Emílio Blum, 131, Sala 409, Bloco A – Centro – Florianópolis – CEP:88020-010 – Brasil

Tel: +55 48 3879-6888 – email: [contato@hydroinfo.com.br](mailto:contato@hydroinfo.com.br), [jun@hydroinfo.com.br](mailto:jun@hydroinfo.com.br)



## Modernização de ETEs

**Otimização das ETEs:** crucial para melhoria da qualidade e eficiência do tratamento de efluentes.

Desenvolvimento de novos sensores e técnicas de monitoramento = **grande quantidade de dados subutilizados.**

Crescente demanda por **soluções que reduzam custos de operação** e melhorem a qualidade do efluente tratado.

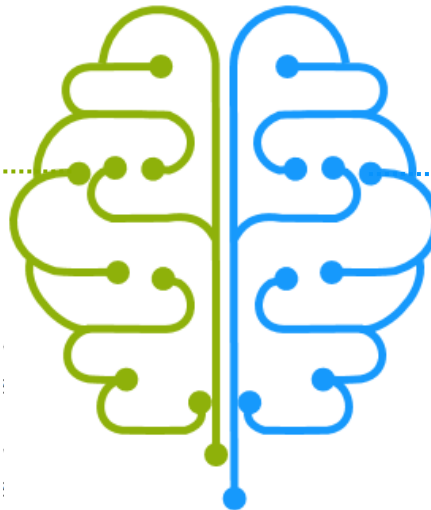


# Aprendizado de Máquina em ETEs: Aplicações e Desafios

## Aplicação do ML em ETEs

ML treinado com dados identifica padrões para aprimorar ETEs.

Uso de diversos algoritmos para previsão de inúmeros parâmetros em diferentes etapas da operação.



## Desafios na aplicação do ML em ETEs

Construção de um banco de dados robusto e unificado.

Variação de dados de diferentes fontes e intervalos de tempo.

Limitações de sensores e anotações qualitativas.

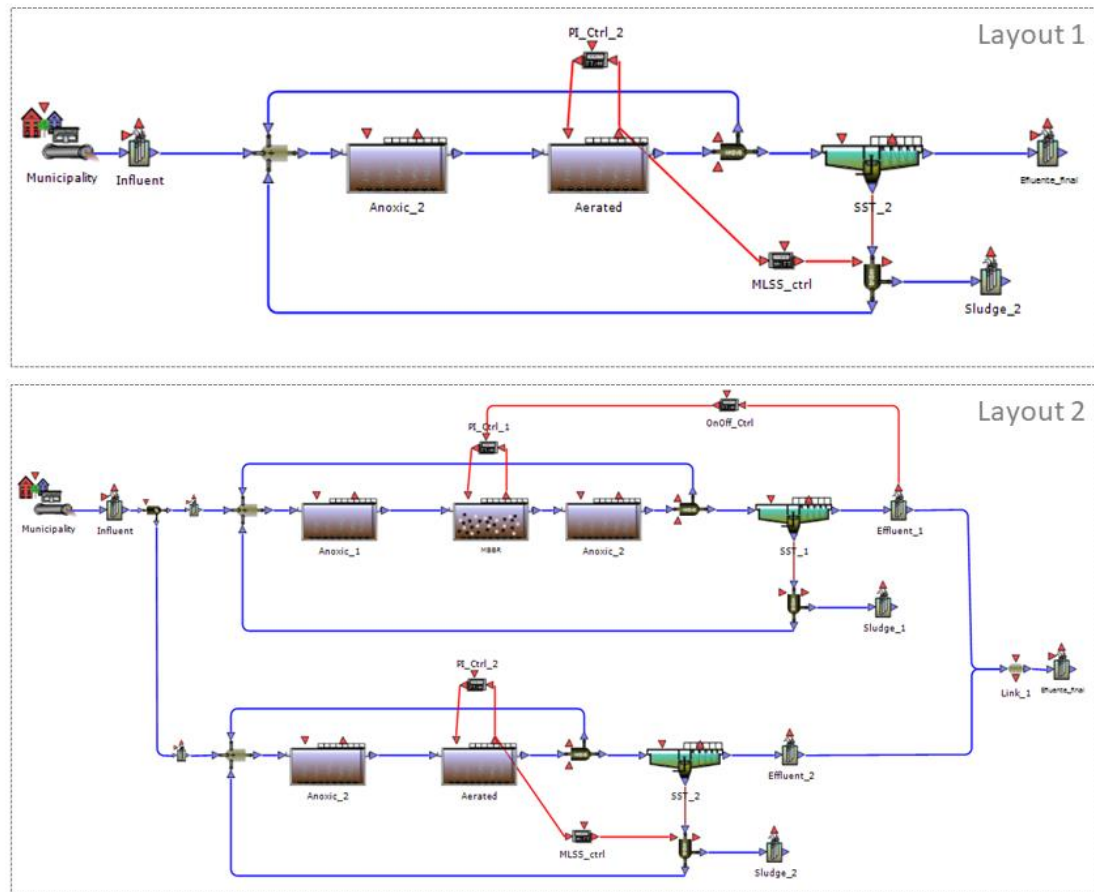
Mudanças frequentes no processo e dificuldades na atribuição de causas.

Coleta, limpeza, discretização e normalização de dados = boas práticas de coleta de dados.

# Objetivo do trabalho

Avaliar o uso de modelos preditivos de ML para previsão de parâmetros de qualidade de efluentes a partir de dados de entrada e saída obtidos em uma ETE simulada (WEST) e em uma ETE real de uma indústria cervejeira (AMBEV Lages/SC).

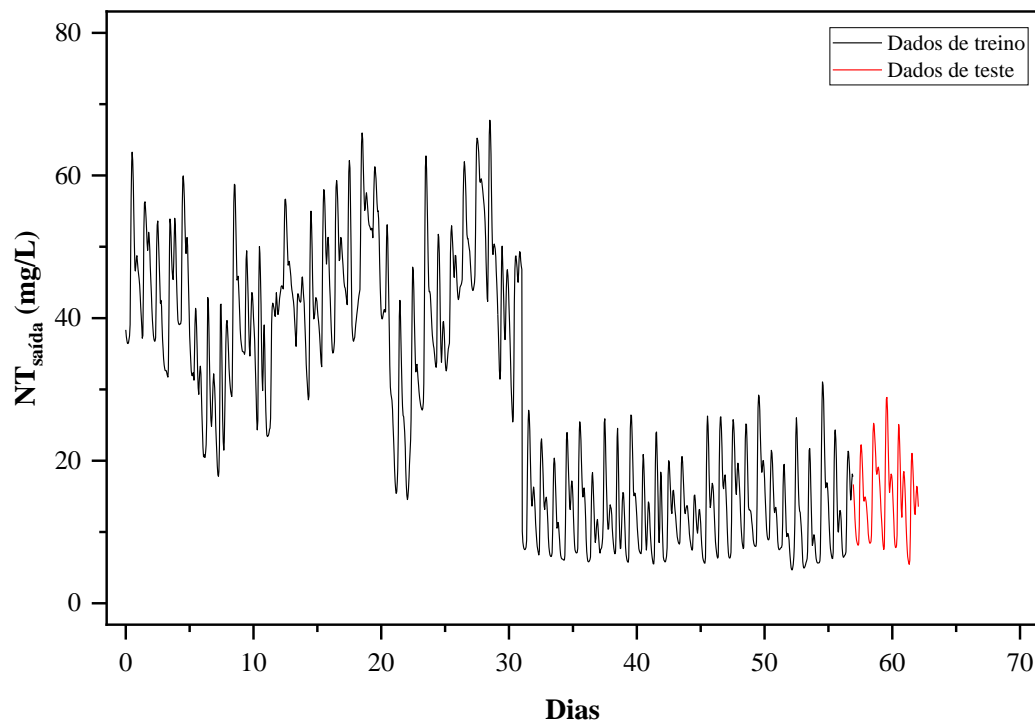
# Estudo de Caso e Descrição do **Conjunto de Dados**



**Figura 1:** Layouts da ETE utilizada para obtenção dos dados no software WEST.

## Dados da ETE simulada (WEST)

- Dois layouts de funcionamento da ETE: Layout 1 e Layout 2.
- 31 dias de operação em cada layout.
- Dados com frequência horária, totalizando 745 dados por variável.
- Variáveis de Processo: DBO, DQO, NH, NO, Q, NKT, **NT** e SST.

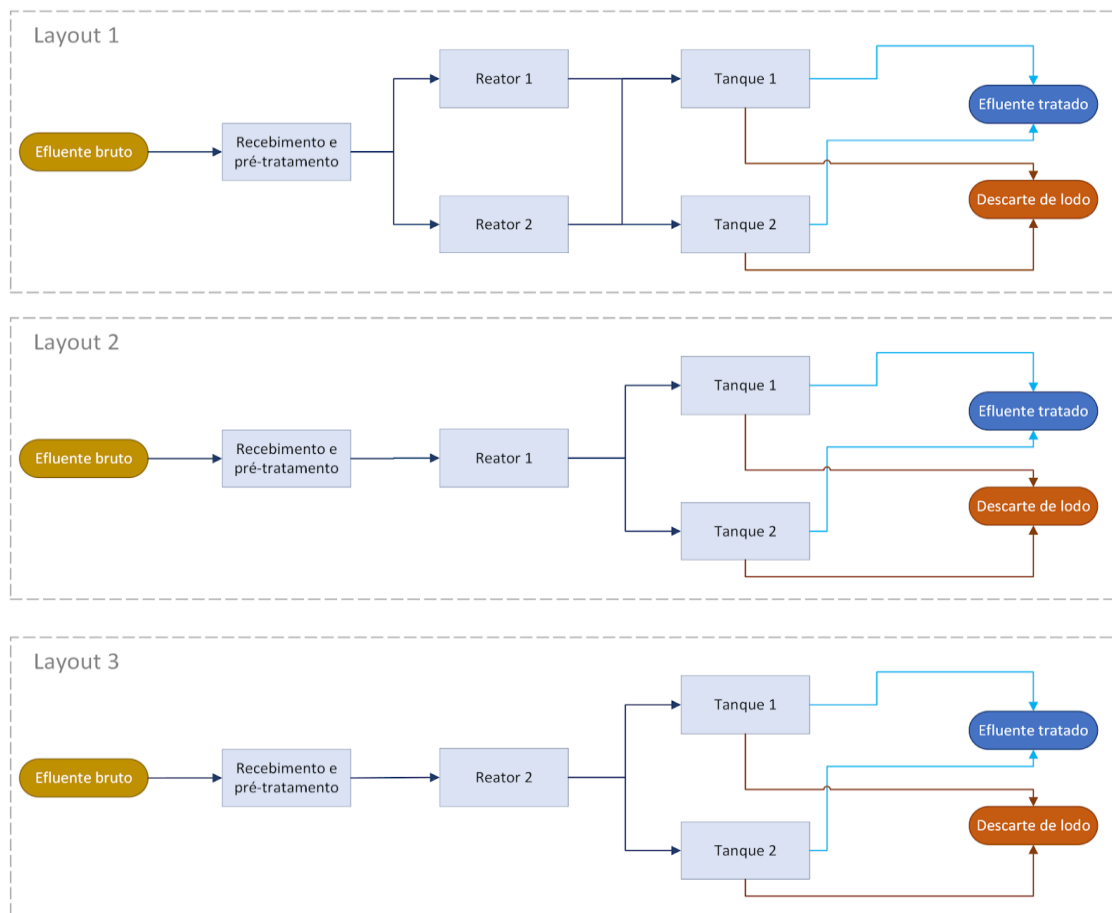


**Figura 2:** Perfil de saída do parâmetro NT para os 62 dias de operação do Layout 1 e Layout 2 da ETE WEST.

## Unificação dos Dados de Simulação

- Dados dos Layouts 1 e 2 unificados.
- Base de dados contínua com 62 dias de operação.
- Perfil de Saída do NT: impacto da inserção da segunda linha de operação.

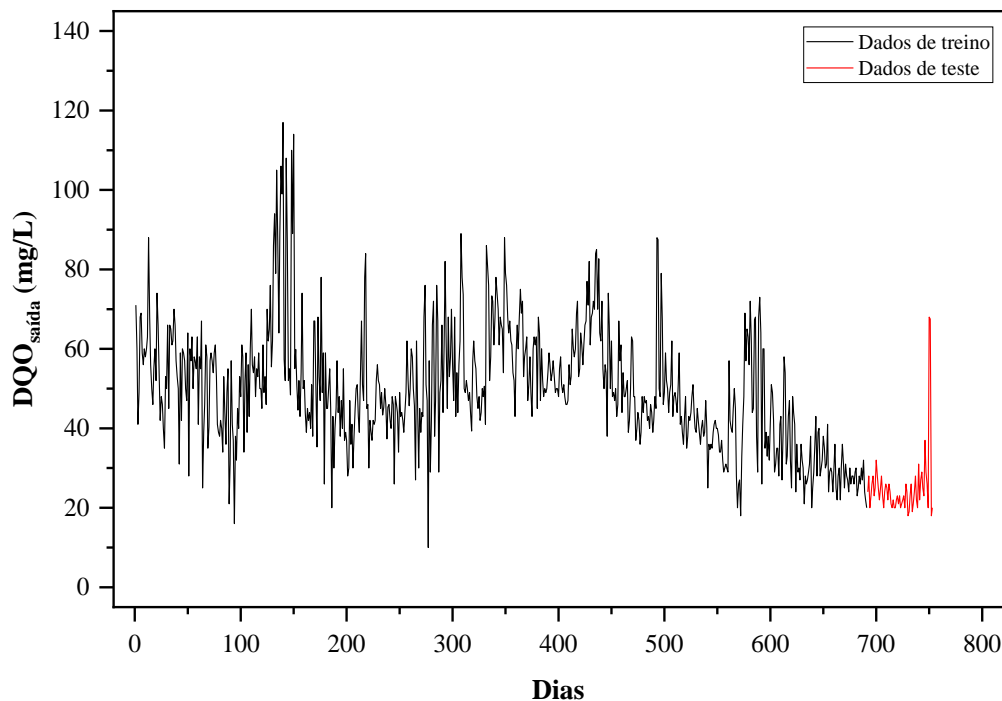
# Estudo de Caso e Descrição do **Conjunto de Dados**



## Dados da ETE AMBEV - Unidade Lages/SC

- Processo anaeróbio seguido de tratamento aeróbio.
- Dados obtidos do sistema de gestão de processos (LiveMES) e de planilhas de controle interno.
- Três layouts simplificados elaborados considerando as informações de funcionamento dos reatores anaeróbios.
- Tempo de residência médio do efluente = 6,8 dias.

**Figura 3:** Fluxograma simplificado da ETE AMBEV – Unidade Lages.



**Figura 4:** Perfil de saída do parâmetro DQO para a ETE AMBEV.

## Banco de dados utilizado

- Parâmetros Q, DQO, CO, T e pH na entrada e saída da ETE.
- Período de jan/2021 a jan/2023 = Total de 752 dias de dados.

## Dados de Saída do DQO

- Tendência decrescente ao longo dos dois anos.
- Redução significativa no último ano = Reflexo de modificações e melhorias na operação.



## Análise Exploratória de Dados

Resumir e examinar os dados previamente à aplicação de qualquer técnica estatística.

Aplicada à ETE WEST e ETE Ambev.

Análises univariadas e bivariadas.

Uso de bibliotecas Python: pandas, pandas-profiling e sweetviz.

## Detecção de Outliers e Dados Anômalos

Condição de  $3\sigma$  em relação à média.

Exclusão de dados que não atendem ao critério.

Identificação de informações redundantes.

## Imputação de Dados

Uso da técnica KNN Imputation para minimizar exclusões.

Manutenção da continuidade dos registros.

## *Feature Engineering*

Aumento da informação para modelos de ML.

ETE WEST: Adição de variável categórica para mudança na operação.

ETE Ambev: Inclusão de colunas para diferentes modos de operação.

Retrocesso temporal para alinhar dados de entrada e saída.

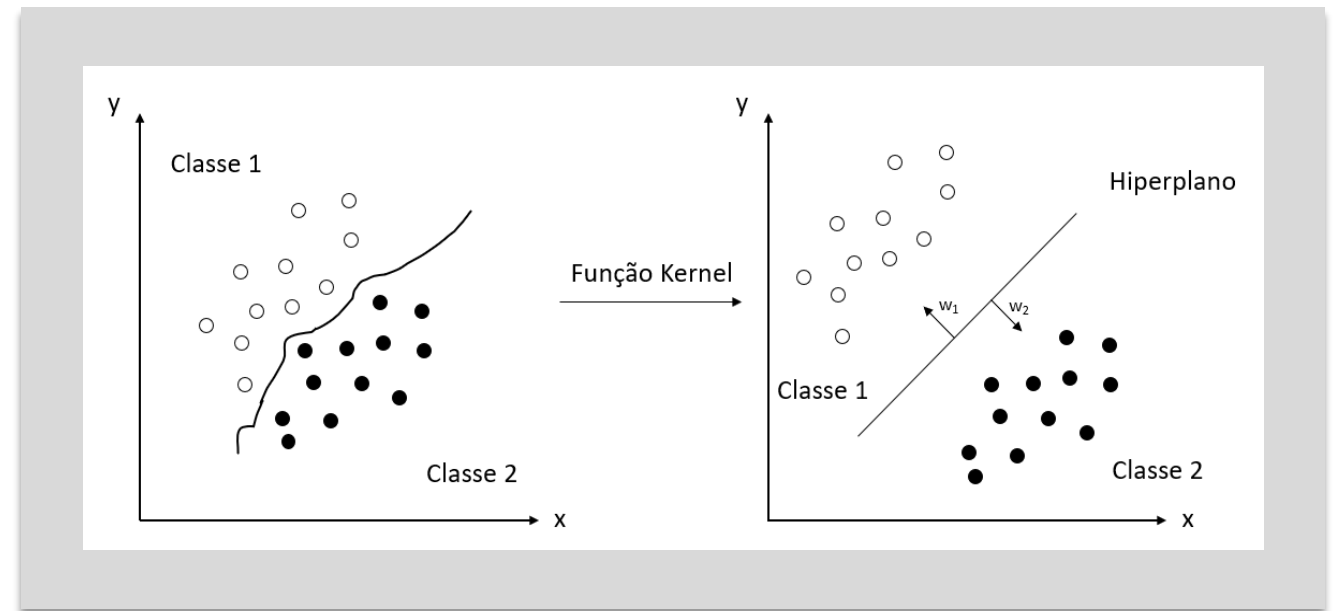
## Máquinas de Vetores de Suporte (SVM)

*Baseado em minimização de risco estrutural.*

*Utilizado para classificação e regressão.*

*Uso de vetores de suporte.*

*Funções kernel: linear, polinomial, RBF, sigmoid.*



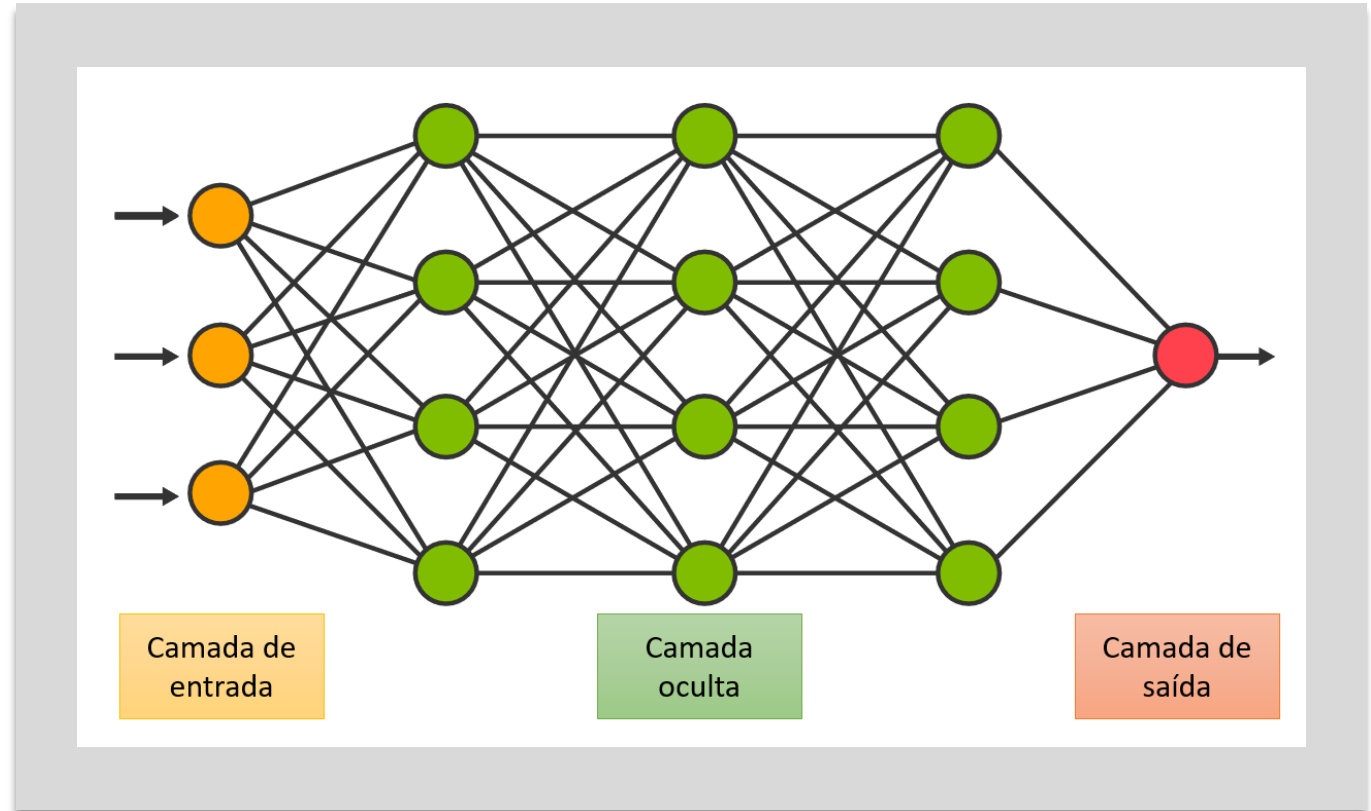
## Perceptron Multicamadas (MLP)

*Rede neural feedforward.*

*Camadas: entrada, saída e ocultas.*

*Treinamento com retropropagação.*

*Função de ativação: ReLU.*



# Modelos de Machine Learning

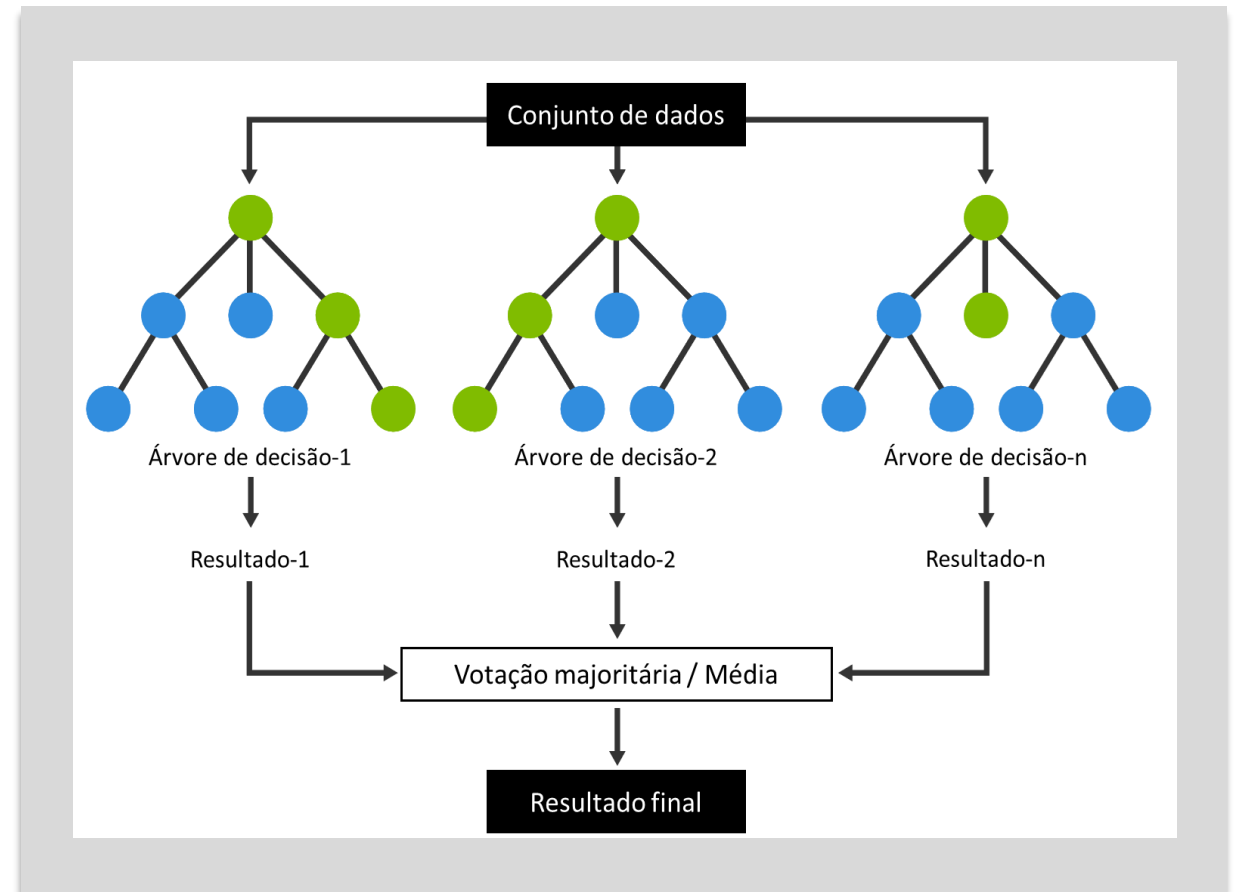
## *Floresta Aleatória (RF)*

*Combina múltiplas árvores de decisão.*

*Utilizado para classificação e regressão.*

*Importância de variáveis.*

*Critério de aprendizado: erro absoluto médio.*



# Cenários de Previsão – ETE WEST

## Cenário 1:

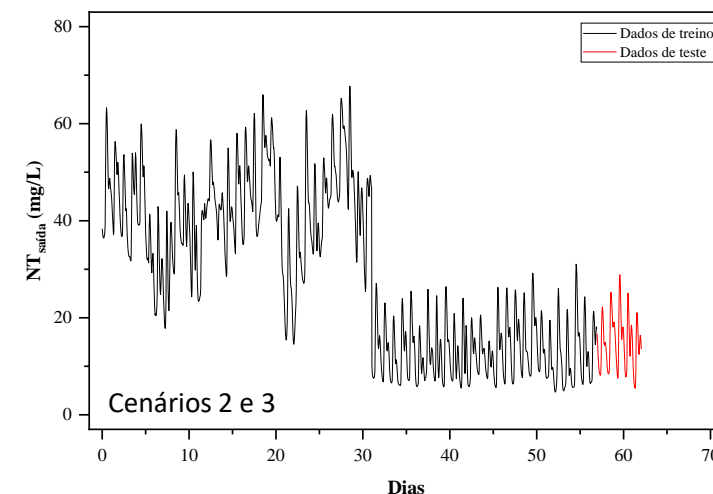
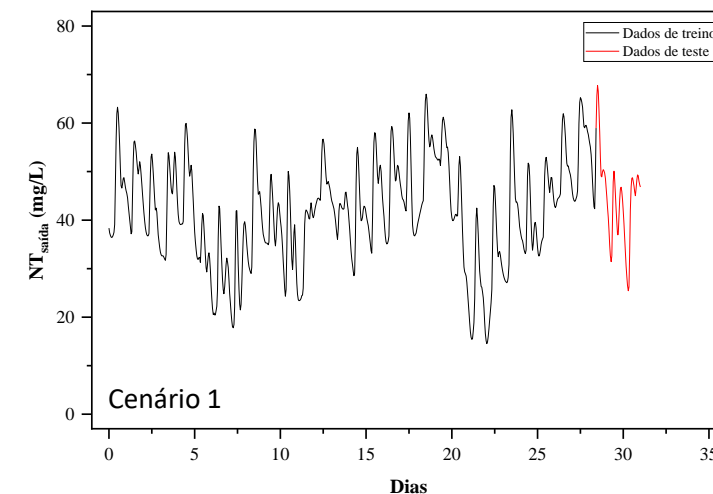
- somente dados do Layout 1 (31 dias de operação)
- 745 registros horários de entrada de todas as variáveis (683 treinamento + 62 teste)

## Cenário 2:

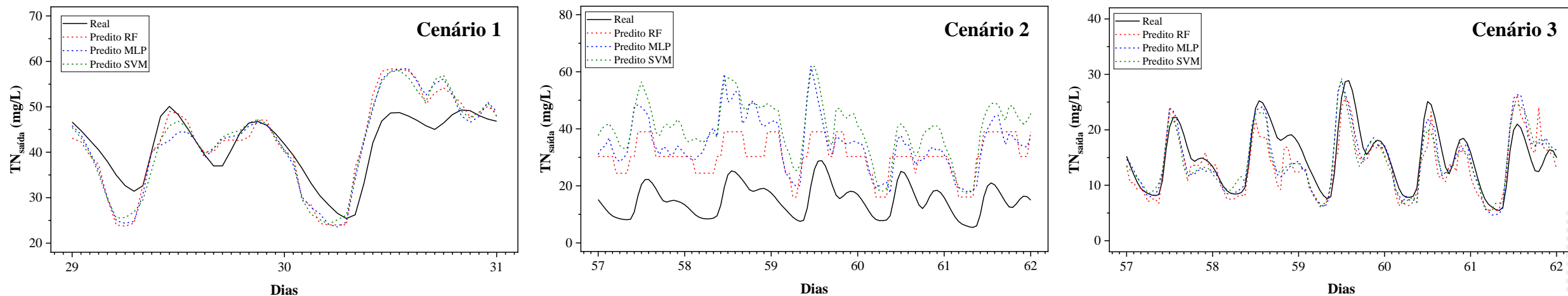
- dados do Layout 1 e Layout 2 (31 + 31 = 62 dias de operação)
- SEM informação adicional sobre o modo de operação da ETE
- 1490 registros horários (1367 treinamento + 123 teste)

## Cenário 3:

- mesma base de dados do cenário 2, porém COM informação adicional sobre o modo de operação da ETE



# Cenários de Previsão – ETE WEST

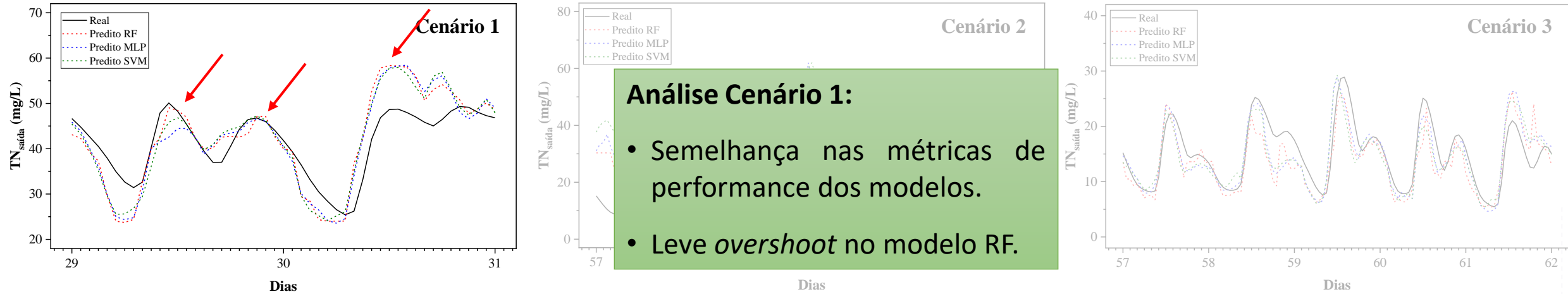


**Figura 5:** Resultados dos modelos preditivos para os dados de NT da ETE WEST usando RF, MLP e SVM.

**Tabela 1:** Comparação de métricas de performance para os modelos preditivos de ML para NT aplicados a ETE WEST.

	Cenário 1			Cenário 2			Cenário 3		
	SVM	MLP	RF	SVM	MLP	RF	SVM	MLP	RF
<b>RMSE</b>	5,76	5,75	5,87	25,88	21,03	16,28	3,18	2,87	3,39
<b>MAE</b>	4,52	4,48	4,61	24,96	19,85	15,58	2,49	2,21	2,55
<b>MAPE</b>	0,11	0,11	0,11	1,85	1,5	1,2	0,16	0,14	0,16
<b>R<sup>2</sup></b>	0,6	0,6	0,58	-21,44	-13,82	-7,89	0,66	0,72	0,61

# Cenários de Previsão – ETE WEST



**Figura 5:** Resultados dos modelos preditivos para os dados de NT da ETE WEST usando RF, MLP e SVM.

**Tabela 1:** Comparação de métricas de performance para os modelos preditivos de ML para NT aplicados a ETE WEST.

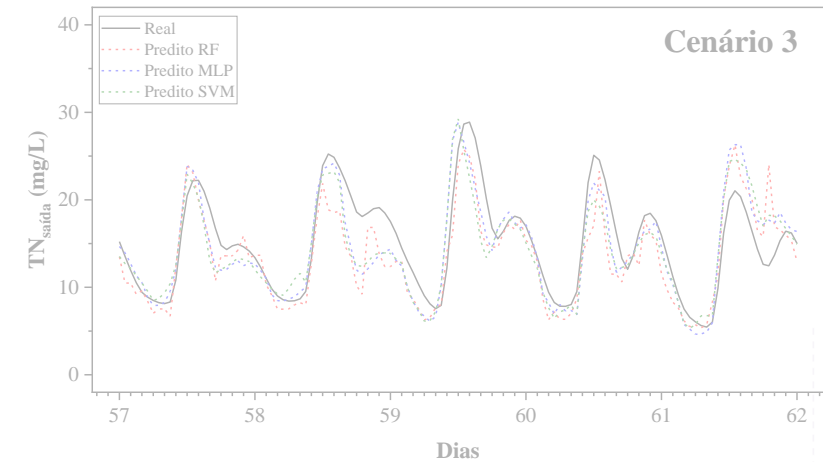
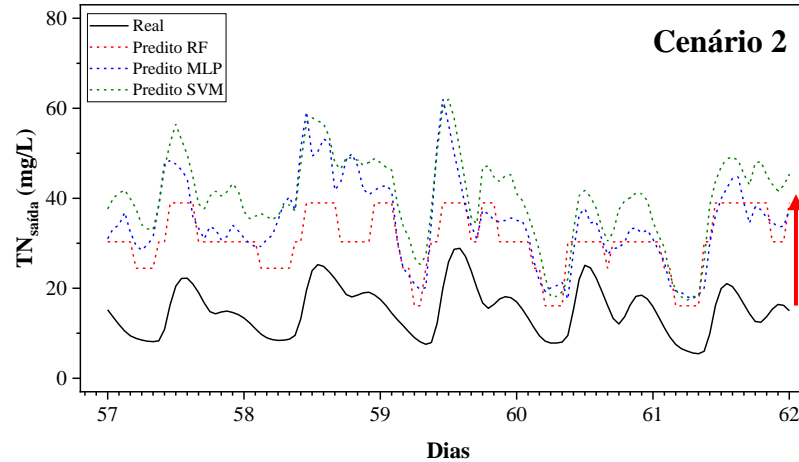
	Cenário 1			Cenário 2			Cenário 3		
	SVM	MLP	RF	SVM	MLP	RF	SVM	MLP	RF
<b>RMSE</b>	5,76	5,75	5,87	25,88	21,03	16,28	3,18	2,87	3,39
<b>MAE</b>	4,52	4,48	4,61	24,96	19,85	15,58	2,49	2,21	2,55
<b>MAPE</b>	0,11	0,11	0,11	1,85	1,5	1,2	0,16	0,14	0,16
<b>R<sup>2</sup></b>	0,6	0,6	0,58	-21,44	-13,82	-7,89	0,66	0,72	0,61



# Cenários de Previsão – ETE WEST

## Análise Cenário 2:

- Modelos com perda de qualidade em comparação com o Cenário 1.
- Modelos não identificaram mudança no funcionamento da ETE.
- Todos os modelos previram valores mais elevados (condizentes com a condição de TN do Cenário 1).



para os dados de NT da ETE WEST usando RF, MLP e SVM.

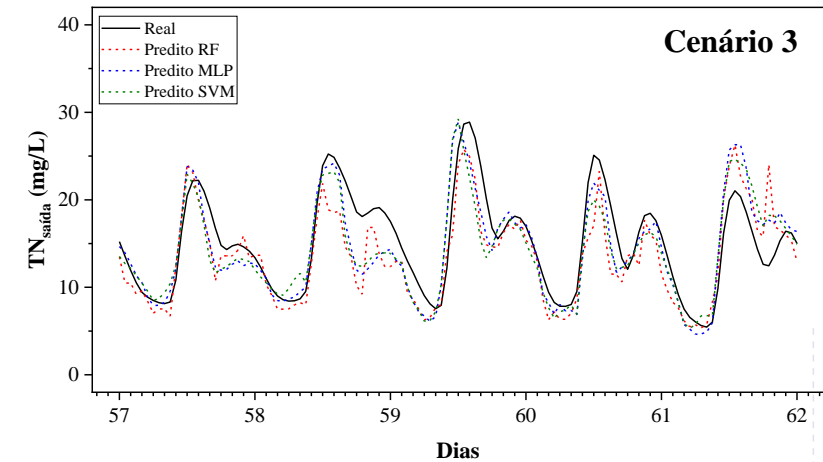
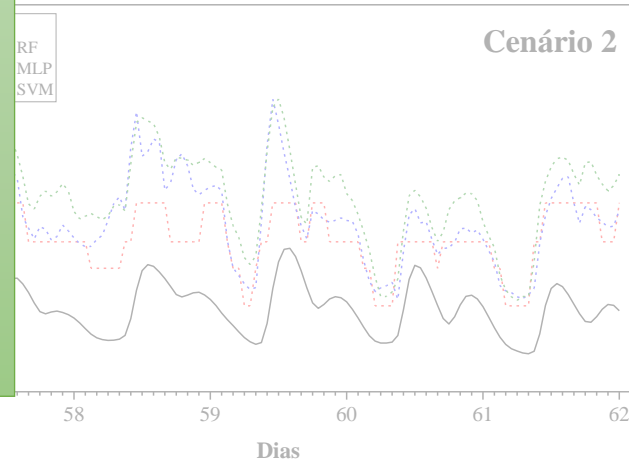
**Tabela 1:** Comparação de métricas de performance para os modelos preditivos de ML para NT aplicados a ETE WEST.

	Cenário 1			Cenário 2			Cenário 3		
	SVM	MLP	RF	SVM	MLP	RF	SVM	MLP	RF
<b>RMSE</b>	5,76	5,75	5,87	25,88	21,03	16,28	3,18	2,87	3,39
<b>MAE</b>	4,52	4,48	4,61	24,96	19,85	15,58	2,49	2,21	2,55
<b>MAPE</b>	0,11	0,11	0,11	1,85	1,5	1,2	0,16	0,14	0,16
<b>R<sup>2</sup></b>	0,6	0,6	0,58	-21,44	-13,82	-7,89	0,66	0,72	0,61

# Cenários de Previsão – ETE WEST

## Análise Cenário 3

- Melhora nas métricas de performance.
- Modelo MLP apresentou as melhores métricas.
- Resultados indicam uma boa aproximação da relação entre dados de entrada e NT de saída.

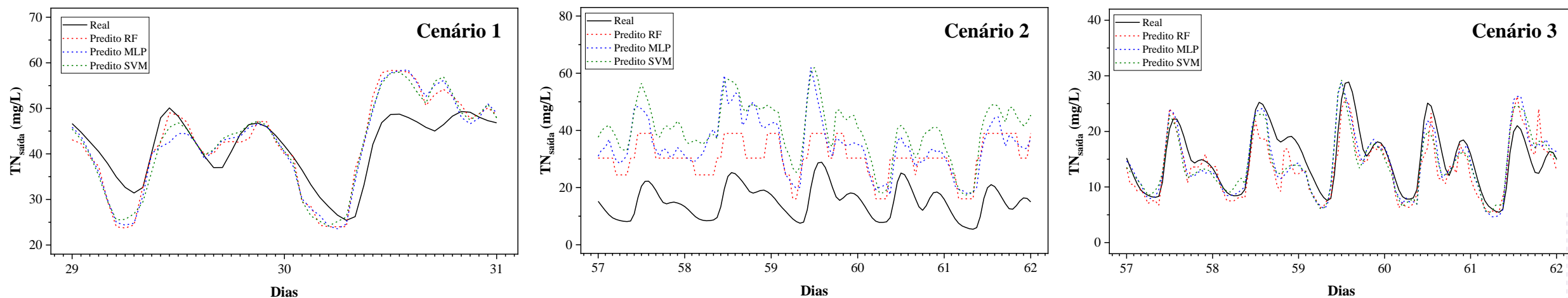


**Figura 5:** Resultados dos modelos preditivos para os dados de NT da ETE WEST usando RF, MLP e SVM.

**Tabela 1:** Comparação de métricas de performance para os modelos preditivos de ML para NT aplicados a ETE WEST.

	Cenário 1			Cenário 2			Cenário 3		
	SVM	MLP	RF	SVM	MLP	RF	SVM	MLP	RF
<b>RMSE</b>	5,76	5,75	5,87	25,88	21,03	16,28	3,18	2,87	3,39
<b>MAE</b>	4,52	4,48	4,61	24,96	19,85	15,58	2,49	2,21	2,55
<b>MAPE</b>	0,11	0,11	0,11	1,85	1,5	1,2	0,16	0,14	0,16
<b>R<sup>2</sup></b>	0,6	0,6	0,58	-21,44	-13,82	-7,89	0,66	0,72	0,61

# Cenários de Previsão – ETE WEST



**Figura 5:** Resultados dos modelos preditivos para os dados de NT da ETE WEST usando RF, MLP e SVM.

Os modelos preditivos, especialmente para o Cenário 3 que inclui a informação categórica de funcionamento da ETE, demonstram potencial para prever o NT de saída da ETE WEST, com métricas de performance favoráveis.

# Cenários de Previsão – ETE AMBEV

## Cenário 1:

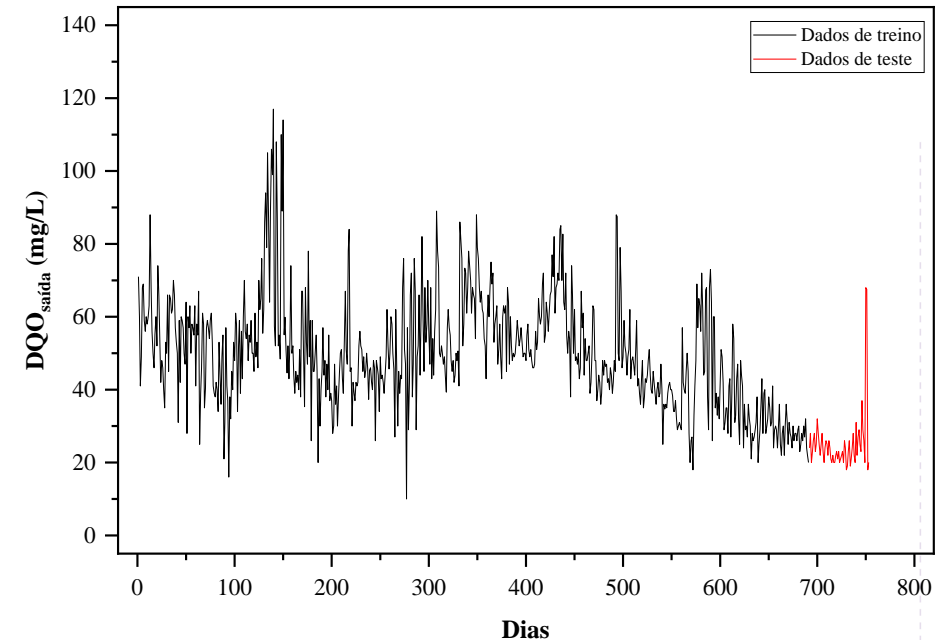
- SEM informação sobre o modo de operação da ETE
- 752 registros diários (690 treinamento + 62 teste)

## Cenário 2:

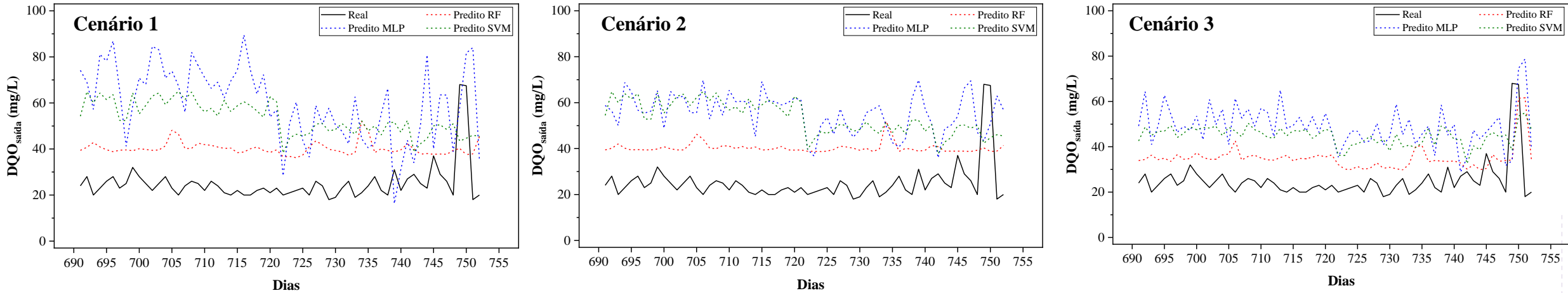
- dados cenário 1 e COM informação sobre modo de operação da ETE (operação dos Reatores 1 e 2).

## Cenário 3:

- dados cenário 2 + dados de Q, DQO, CO, T e pH de saída da ETE do dia anterior foram adicionados como variáveis independentes.



# Cenários de Previsão – ETE AMBEV

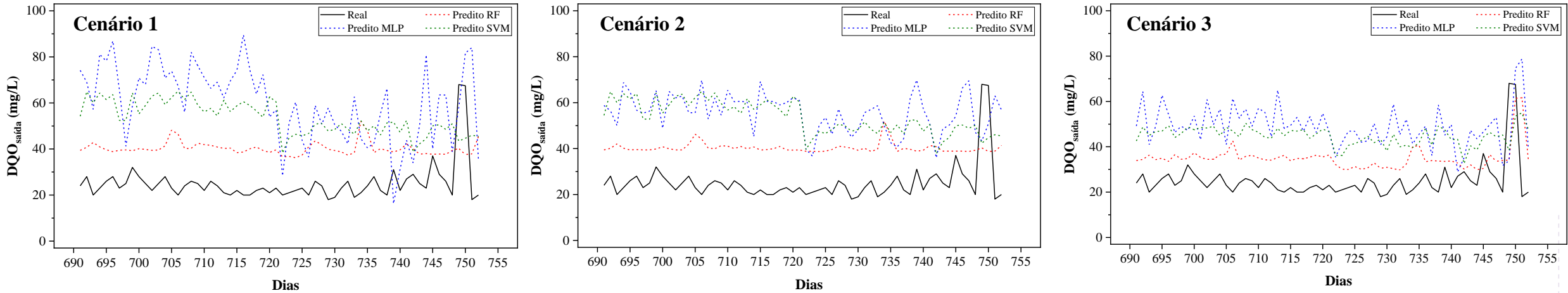


**Figura 5:** Resultados dos modelos preditivos para os dados de DQO da ETE AMBEV usando RF, MLP e SVM.

**Tabela 1:** Comparação de métricas de performance para os modelos preditivos de ML para DQO aplicados a ETE AMBEV.

	Cenário 1			Cenário 2			Cenário 3		
	SVM	MLP	RF	SVM	MLP	RF	SVM	MLP	RF
<b>RMSE</b>	30,86	39,47	17,77	30,95	33,00	17,66	21,74	26,83	13,09
<b>MAE</b>	29,94	35,84	16,97	30,03	31,80	16,98	21,10	25,07	11,37
<b>MAPE</b>	1,28	1,56	0,73	1,29	1,37	0,73	0,91	1,08	0,50
<b>R<sup>2</sup></b>	-11,97	-20,21	-3,30	-12,04	-13,83	-3,24	-5,44	-8,80	-1,33

# Cenários de Previsão – ETE AMBEV

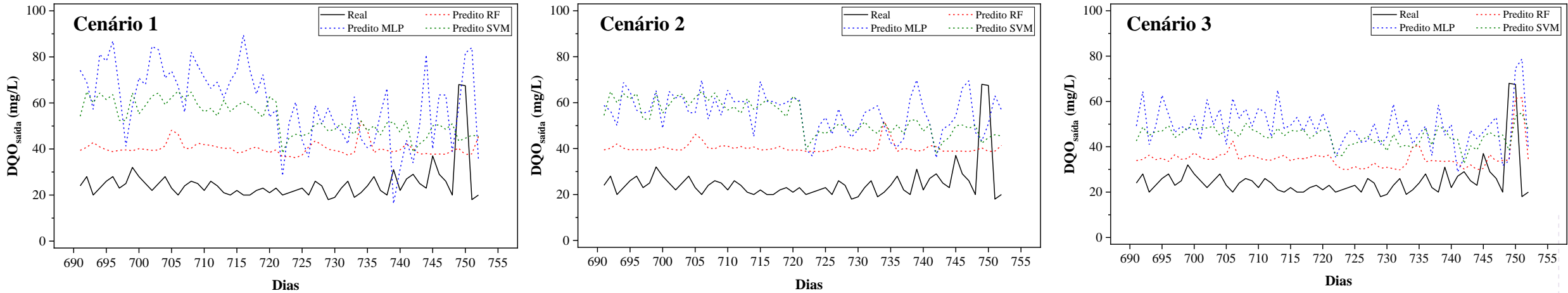


**Figura 6:** Resultados dos modelos preditivos para os dados de DQO da ETE AMBEV usando RF, MLP e SVM.

Tabela 2: C **Análise dos cenários:**

- Tendências foram representadas pelo modelo RF, mas com erros consideráveis.
- A inclusão de outros parâmetros e informações sobre operação pode melhorar os resultados obtidos
- Para essa situação, os modelos não são adequados para tomada de decisões, mas podem ser aprimorados para auxiliar em processos futuros.

# Cenários de Previsão – ETE AMBEV



**Figura 6:** Resultados dos modelos preditivos para os dados de DQO da ETE AMBEV usando RF, MLP e SVM.

Os modelos preditivos, especialmente o modelo RF, mostraram-se promissores, mas requerem mais informações para aprimorar a precisão. A inclusão de variáveis adicionais e dados de operação pode aprimorar esses modelos no futuro.

# Conclusões

Modelos preditivos SVM, MLP e RF foram capazes de prever parâmetros de saída de ETEs.

**ETEs simulada (WEST):** Melhores resultados no cenário 3 com *feature engineering* e modelo MLP.

**ETEs real (AMBEV):** Modelo RF apresentou as melhores métricas. Maior número de informações e variáveis preditivas resultou em melhorias visíveis.

**Feature Engineering:** Destaque para informações relacionadas ao modo de operação e variáveis de processo, que ajudam os modelos a se adaptarem rapidamente a novos cenários.

**Desafios para aplicação de modelos de ML em ETEs:** Necessidade de medição e anotação mais abrangente e frequente dos parâmetros de qualidade e das possíveis alterações nas operações de ETEs.

Em condições controladas e sem mudanças abruptas, modelos de ML podem prever valores e tendências com base em dados de entrada.



# Agradecimentos

- Conselho Nacional de Pesquisa e Desenvolvimento – CNPq e Ministério da Ciência, Tecnologia e Inovações pelo apoio financeiro e pelas bolsas de fomento tecnológico (Processos nº 350509/2022-0, 350508/2022-4, 424532/2021-2).
- Hydroinfo e DHI.
- LABMAC/UFSC.
- AMBEV Unidade Lages/SC.

