

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

RICARDO EMANUEL VAZ VARGAS

**BASE DE DADOS E *BENCHMARKS* PARA PROGNÓSTICO DE
ANOMALIAS EM SISTEMAS DE ELEVAÇÃO DE PETRÓLEO**

VITÓRIA
2019

RICARDO EMANUEL VAZ VARGAS

**BASE DE DADOS E *BENCHMARKS* PARA PROGNÓSTICO DE
ANOMALIAS EM SISTEMAS DE ELEVAÇÃO DE PETRÓLEO**

Tese submetida ao Programa de Pós-Graduação
em Engenharia Elétrica do Centro Tecnológico
da Universidade Federal do Espírito Santo,
como requisito parcial para obtenção do grau
acadêmico de Doutor em Engenharia Elétrica -
Automação.

Orientador: Prof. Dr. Evandro Ottoni Teatini
Salles

VITÓRIA
2019

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

Vargas, Ricardo Emanuel Vaz, 1981-
V297b Base de dados e benchmarks para prognóstico de anomalias
em sistemas de elevação de petróleo / Ricardo Emanuel Vaz
Vargas. - 2019.
74 f. : il.

Orientador: Evandro Ottoni Teatini Salles.
Tese (Doutorado em Engenharia Elétrica) - Universidade
Federal do Espírito Santo, Centro Tecnológico.

1. Detecção e diagnóstico de falhas. 2. Monitoramento de poços
de petróleo. 3. Gerenciamento de eventos anormais. 4.
Classificação de séries temporais multivariáveis. I. Salles,
Evandro Ottoni Teatini. II. Universidade Federal do Espírito
Santo. Centro Tecnológico. III. Título.

CDU: 621.3

RICARDO EMANUEL VAZ VARGAS

BASE DE DADOS E BENCHMARKS PARA PROGNÓSTICO DE ANOMALIAS EM SISTEMAS DE ELEVAÇÃO DE PETRÓLEO

Tese submetida ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau acadêmico de Doutor em Engenharia Elétrica – Automação.

Aprovada em 27 de agosto de 2019.

COMISSÃO EXAMINADORA

Evandro Ottoni Teatini Salles

Prof. Dr. Evandro Ottoni Teatini Salles
Universidade Federal do Espírito Santo
Orientador

Leizer Schnitman

Prof. Dr. Leizer Schnitman
Universidade Federal da Bahia

Luis Martí Orosa

Prof. Dr. Luis Martí Orosa
Universidade Federal Fluminense

Mario Cesar Massa de Campos

Dr. Mario Cesar Massa de Campos
Petróleo Brasileiro S/A - CENPES

Thomas Walter Rauber

Prof. Dr. Thomas Walter Rauber
Universidade Federal do Espírito Santo

VITÓRIA
2019

AGRADECIMENTOS

Tenho muitas pessoas a agradecer!

Aos meus pais, pela vida e por terem me oferecido todo suporte necessário até então.

À minha esposa, pelo incentivo de sempre, pela parceria na vida e por compreender os grandes sacrifícios exigidos por um doutorado, especialmente quando feito em conjunto com trabalho em empresa. Ao meu filho Davi, pela sua alegria contagiante, que me anima bastante.

À Maria José, por cuidar diariamente da minha família há mais de 14 anos.

Ao Prof. Dr. Celso José Munaro e ao Prof. Dr. Patrick Marques Ciarelli, pela grande dedicação a este trabalho e publicações relacionadas a ele, inclusive em vários momentos fora do horário de expediente, pela paciência e foco durante as extensas reuniões, pela flexibilidade para alinhamento de agendas e pelas valiosas sugestões de melhoria. Profissionais assim transformam sociedades.

À Petrobras, por ter permitido as publicações dos artigos desenvolvidos neste doutorado e por ter liberado 20% da minha carga-horária de trabalho para a realização das atividades desta pós-graduação, do seu início ao fim. Foi investida uma carga-horária muito maior que essa ao longo desse período, mas sem esse incentivo formal esta pesquisa nem teria começado.

Ao meu ex-gerente Heber Barbosa, pelo incentivo inicial sem o qual a minha liberação parcial na Petrobras não teria acontecido. Esse nasceu para ser gerente, e dos bons!

Aos colegas de trabalho na Petrobras que me prestaram apoio ao longo dessa jornada, em particular: Bento Daher, Bruno Amaral, Carlos Gastão, Cézar Siqueira, Edson Gonoring, Helder Fábio, Jorge Lins, Lucas Magalhães, Luiz Carlos, Rafael Borin, Rogério Leite e Rômulo Furtado. Sem esse apoio, os resultados deste trabalho teriam sido muito diferentes. Agradeço especialmente ao Mario Campos, por ter sido o meu orientador no âmbito da Petrobras ao longo desta pesquisa. Sem as suas sugestões, talvez eu teria seguido outro rumo menos proveitoso. Agradeço também de forma especial ao André Medeiros, ao Daniel Centurion e ao Jean Araújo, pelas produtivas parcerias, pelas verdadeiras aulas sobre o simulador OLGA e pelas consultorias na área Elevação e Escoamento de Petróleo.

Realmente obrigado!

VITÓRIA
2019

In God we trust, all others bring data.

William Edwards Deming (1900-1993)

VITÓRIA
2019

RESUMO

A indústria de petróleo considera que prognóstico de anomalias em poços produtores de petróleo pode ajudar a reduzir custos de manutenção e a evitar perdas de produção e acidentes ambientais e à vida humana. Um poço de produção de petróleo se refere a um conjunto de sensores e de sistemas mecânicos, elétricos e hidráulicos. Como em praticamente qualquer processo industrial, também ocorrem variados tipos de anomalias em processo de elevação e escoamento de petróleo. Esta tese formula e avalia as hipóteses que anomalias em sistemas de elevação natural de petróleo podem ser detectadas com Aprendizado de Máquina e que o uso de instâncias simuladas e desenhadas à mão por especialistas é uma solução viável para treinamento de detectores de anomalias reais raras. A escassez de medições nesse tipo de processo, devido à baixa confiabilidade da instrumentação em tais ambientes hostis, é uma desvantagem. Outra questão é a ausência de dados de anomalias – em quantidade, em qualidade e adequadamente estruturados – em sistemas de elevação natural de petróleo. Para contribuir com abordagens para prognóstico desse tipo de anomalia baseadas em Aprendizado de Máquina, este trabalho preparou e tornou público um conjunto de dados original e realista com instâncias de oito tipos de anomalias caracterizadas por oito variáveis de processo. Muitas horas de trabalho conjunto com engenheiros da Petróleo Brasileiro S.A. especialistas na área Elevação e Escoamento de Petróleo foram necessárias para validar instâncias históricas e para produzir instâncias simuladas e desenhadas à mão. A metodologia desenvolvida e utilizada nessa preparação é detalhada. Desafios específicos que pesquisadores podem explorar com o conjunto de dados publicado são definidos. Resultados experimentais relacionados a esses desafios sugerem que as hipóteses formuladas são verdadeiras. Este trabalho resultou em duas contribuições relevantes. Um conjunto de dados público e desafiador que pode ser utilizado como referência para desenvolvimento de (i) técnicas de Aprendizado de Máquina para tratamento de dificuldades inerentes a dados reais e (ii) métodos para tarefas específicas associadas a classificação de anomalias em sistemas de elevação natural de petróleo. A outra contribuição são os desafios específicos que foram propostos.

Palavras-chave: Detecção e diagnóstico de falhas. Monitoramento de poços de petróleo. Gerenciamento de eventos anormais. Classificação de séries temporais multivariáveis.

ABSTRACT

The oil industry considers that prognosis of anomalies in oil-producing wells can help to reduce production losses, environmental accidents, and human casualties and reduce maintenance costs. An oil well refers to a set of sensors and mechanical, pneumatic, and hydraulic systems. As in virtually any industrial process, several types of anomalies also occur in the process of oil lifting and flow assurance. This thesis formulates and evaluates the hypothesis that anomalies in naturally flowing wells can be detected with Machine Learning and that the use of expert hand-drawn and simulated instances is a feasible solution for the training of rare actual anomalies' detectors. The scarcity of measurements in such processes is a drawback due to the low reliability of instrumentation in such hostile environments. Another issue is the absence of anomalies' data – in quantity, quality, and adequately structured – in naturally flowing wells. To contribute to Machine Learning-based approaches to the prognosis of this type of anomaly, this work prepared and made public an original and realistic dataset with instances of eight types of anomalies characterized by eight process variables. Many hours of working together with experts from Petróleo Brasileiro S.A. were required to validate historical instances and to produce simulated and hand-drawn instances. The methodology developed and used in this preparation is detailed. Specific challenges that researchers can explore with the published dataset are defined. Experimental results related to these challenges suggest that the formulated hypotheses are true. This work has resulted in two relevant contributions. A challenging public dataset that can be used as a benchmark for the development of (i) machine learning techniques related to inherent difficulties of actual data, and (ii) methods for specific tasks associated with anomalies' classification in naturally flowing wells. The other contribution is the proposal of the defined benchmarks.

Keywords: Fault detection and diagnosis. Oil well monitoring. Abnormal event management. Multivariate time series classification.

LISTA DE FIGURAS

Figura 1 – Ilustração das quatro etapas do transporte de petróleo.....	14
Figura 2 – Esquema de um poço marítimo surgente de petróleo.....	29
Figura 3 – Gráficos relacionados a um exemplo de ocorrência real de aumento abrupto de BSW. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.	32
Figura 4 – Gráficos relacionados a um exemplo de ocorrência real de fechamento espúrio de DHSV. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.	33
Figura 5 – Gráficos relacionados a um exemplo de ocorrência real de intermitência severa. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.....	34
Figura 6 – Gráficos relacionados a um exemplo de ocorrência real de instabilidade no fluxo. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.....	35
Figura 7 – Gráficos relacionados a um exemplo de ocorrência real de perda rápida de produtividade. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.....	36
Figura 8 – Gráficos relacionados a um exemplo de ocorrência real de restrição em CKP. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.....	37
Figura 9 – Gráficos relacionados a um exemplo de ocorrência real de incrustação em CKP. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.....	38
Figura 10 – Gráficos relacionados a um exemplo de ocorrência real de hidrato em linha de produção. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.	39
Figura 11 – Hierarquia de aprendizado utilizada em (FACELI <i>et al.</i> , 2011).	40
Figura 12 – Instância real histórica rotulada como normal. P-TPT e P-MON-CKP em Pa. T-TPT em °C.	50
Figura 13 – Instância real histórica rotulada como anômala. P-TPT e P-MON-CKP em Pa. T-TPT em °C.	50
Figura 14 – Ilustração do modelo de gráfico preenchido por um especialista na anomalia 9.	51
Figura 15 – Mapa de dispersão das instâncias reais históricas do 3W dataset.	56
Figura 16 – Dispersão da F1, por cenário, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.....	60
Figura 17 – Dispersão da precisão, por cenário, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.....	60
Figura 18 – Dispersão da sensibilidade, por cenário, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.	60
Figura 19 – Dispersão da F1, por algoritmo, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.....	61
Figura 20 – Dispersão da precisão, por algoritmo, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.	61
Figura 21 – Dispersão da sensibilidade, por algoritmo, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.	61
Figura 22 – Dispersão da F1 nos resultados da experimentação relacionada a detecção de anomalia.	64
Figura 23 – Dispersão da precisão nos resultados da experimentação relacionada a detecção de anomalia.	64
Figura 24 – Dispersão da sensibilidade nos resultados da experimentação relacionada a detecção de anomalia.	65

LISTA DE TABELAS

Tabela 1 – Estimativas de tamanhos de janelas temporais utilizados para confirmar ocorrências de anomalias.....	39
Tabela 2 – Quantidades de instâncias que compõem o 3W dataset.	56
Tabela 3 – Médias e desvio padrão das métricas consideradas, por cenário, na experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas. Foram obtidos valores iguais para as três métricas devido ao uso do parâmetro “ <i>average='micro'</i> ” (SCIKIT-LEARN, 2019).	59
Tabela 4 – Médias e desvio padrão das métricas consideradas, por algoritmo, na experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas. Foram obtidos valores iguais para as três métricas devido ao uso do parâmetro “ <i>average='micro'</i> ” (SCIKIT-LEARN, 2019).	60
Tabela 5 – Resultados do Teste de Holm no contexto dos cenários da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.	62
Tabela 6 – Principais resultados das experimentações para o <i>benchmark</i> sobre o impacto do uso de instâncias simuladas e desenhadas à mão.	62
Tabela 7 – Médias e desvio padrão das métricas consideradas, por algoritmo, na experimentação relacionada a detecção de anomalia. Foram obtidos valores iguais para as três métricas devido ao uso do parâmetro “ <i>average='micro'</i> ” (SCIKIT-LEARN, 2019).	64
Tabela 8 – Resultados do Teste de Holm no contexto dos algoritmos da experimentação relacionada a detecção de anomalia.....	65
Tabela 9 – Principais resultados das experimentações para o <i>benchmark</i> sobre detecção de anomalia.....	65

LISTA DE SIGLAS

AEM	– <i>Abnormal Event Management</i>
ANM	– Árvore de Natal Molhada
ANP	– Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
API	– <i>Application Programming Interface</i>
BCS	– Bombeio Centrífugo Submerso
BCSS	– Bombeio Centrífugo Submerso Submarino
BM	– Bombeio Mecânico
BSW	– <i>Basic Sediment and Water</i>
CKP	– <i>Choke de Produção</i>
COI	– Centro de Operações Integradas
CSV	– <i>Comma-Separated Values</i>
DHSV	– <i>Down Hole Safety Valve</i>
DSV	– <i>Downhole Safety Valve</i>
ECG	– Eletrocardiograma
EEG	– Eletroencefalograma
ESP	– <i>Electrical Submersible Pumps</i>
E&P	– Exploração e Produção de Petróleo
F1	– Medida F
FN	– Falso-negativo
FP	– Falso-positivo
GEA	– Gerenciamento de Eventos Anormais
JIP	– <i>Joint Industry Project</i>
MAE	– Monitoramento de Alarmes Especialistas
MTS	– <i>Multivariate Time Series</i>
P	– Precisão
PDG	– <i>Permanent Downhole Gauge</i>
PPGEE	– Programa de Pós-Graduação em Engenharia Elétrica
P-MON-CKP	– Pressão do fluido montante à válvula CKP
P-PDG	– Pressão do fluido no PDG
P-TPT	– Pressão do fluido no TPT
RBF	– <i>Radial Base Function</i>
RIFTS	– <i>Reliability Information and Failure Tracking System</i>
S	– Sensibilidade
STM	– Série Temporal Multivariável
SVM	– <i>Support Vector Machine</i>
TPT	– <i>Temperature and Pressure Transducer</i>
T-JUS-CKP	– Temperatura do fluido jusante à válvula CKP
T-TPT	– Temperatura do fluido no TPT
UEP	– Unidade Estacionária de Produção
UO-ES	– Unidade Operacional da Petrobras no Espírito Santo
VN	– Verdadeiro-negativo
VP	– Verdadeiro-positivo

SUMÁRIO

1	INTRODUÇÃO	12
1.1	CONTEXTUALIZAÇÕES	12
1.1.1	Anomalia e Gerenciamento de Eventos Anormais.....	12
1.1.2	Sistemas de Elevação e Escoamento de Petróleo	13
1.1.3	Estados de um Poço Produtor de Petróleo Operado via Elevação Natural	17
1.1.4	Aprendizado de Máquina e Ciência de Dados	18
1.1.5	Ambiente de Desenvolvimento desta Tese.....	18
1.2	MOTIVAÇÃO.....	19
1.3	HIPÓTESES	20
1.4	OBJETIVOS.....	21
1.5	JUSTIFICATIVAS.....	21
1.6	CONTRIBUIÇÕES	26
1.7	ORGANIZAÇÃO DA TESE	27
2	DESCRIÇÃO DO PROBLEMA	28
2.1	POÇOS MARÍTMOS SURGENTES DE PETRÓLEO	28
2.2	SENSORES GERALMENTE DISPONÍVEIS	29
2.3	ANOMALIAS MAIS CRÍTICAS	30
2.3.1	Aumento Abrupto de BSW	31
2.3.2	Fechamento Espúrio de DHSV	32
2.3.3	Intermitênci a Severa.....	33
2.3.4	Instabilidade no Fluxo	34
2.3.5	Perda Rápida de Produtividade	35
2.3.6	Restrição Rápida em CKP	36
2.3.7	Incrustação em CKP	37
2.3.8	Hidrato em Linha de Produção.....	38
2.4	TAMANHOS DE JANELAS TEMPORAIS DURANTE MONITORAMENTO	39
3	FUNDAMENTAÇÃO TEÓRICA.....	40
3.1	FUNCIONAMENTO DOS ALGORITMOS DE APRENDIZADO DE MÁQUINA.....	40
3.2	SÉRIE TEMPORAL MULTIVARIÁVEL	42
3.3	ALGORITMOS EXPERIMENTADOS.....	43
3.3.1	<i>Decision Tree</i>	44
3.3.2	<i>AdaBoost</i>	44
3.3.3	<i>Random Forest</i>	44
3.3.4	<i>Isolation Forest</i>	45
3.3.5	<i>Gradient Boosting</i>	45
3.3.6	<i>Support Vector Machine</i>	46
3.3.7	<i>One Class SVM</i>	46
3.4	MÉTRICAS DE DESEMPENHO.....	46

4	METODOLOGIA.....	48
4.1	PREPARAÇÃO DO 3W DATASET	48
4.2	<i>BENCHMARKS PROPOSTOS</i>	52
4.2.1	Impacto do Uso de Instâncias Simuladas e Desenhadas à Mão	52
4.2.2	Detecção de Anomalia.....	54
5	RESULTADOS.....	55
5.1	3W DATASET	55
5.2	IMPACTO DO USO DE INSTÂNCIAS SIMULADAS E DESENHADAS À MÃO.....	57
5.3	DETECÇÃO DE ANOMALIA.....	62
6	CONCLUSÕES.....	66
	REFERÊNCIAS.....	68

1 INTRODUÇÃO

Para tornar mais clara esta tese, este capítulo é dividido em seções que descrevem os seus principais aspectos introdutórios: contextualizações, motivação, hipóteses, objetivos, justificativas e contribuições. A última seção deste capítulo resume como o restante deste documento é organizado.

1.1 CONTEXTUALIZAÇÕES

Nesta seção, são feitas cinco contextualizações importantes para o integral entendimento desta tese. A primeira explica os conceitos utilizados de anomalia e de Gerenciamento de Eventos Anormais (GEA). A segunda descreve os principais tipos de sistemas de elevação e escoamento de petróleo que este trabalho pode beneficiar. A terceira diz respeito aos diferentes estados em que um poço produtor de petróleo pode se encontrar. A quarta é uma introdução sobre Aprendizado de Máquina, a área da Ciência que abarca boa parte das técnicas utilizadas neste trabalho. A última explica o ambiente no qual esta tese foi concebida e desenvolvida.

1.1.1 Anomalia e Gerenciamento de Eventos Anormais

Na indústria em geral, tem aumentado as demandas por mais segurança operacional, produtividade, qualidade e eficiência energética (JOUNELA, 2007). Complexidade, instrumentação e automação têm aumentado significativamente nesse contexto para suprir essas demandas (VENKATASUBRAMANIAN *et al.*, 2003). Malhas de controle, sejam manuais ou automatizadas, são desenvolvidas para manterem operações em condições normais, mas há mudanças e distúrbios que esses controles podem não conseguir lidar satisfatoriamente. Falhas ocorrem nessas situações (RUSSELL; CHIANG; BRAATZ, 2000).

O termo “falha” é definido em (RUSSELL; CHIANG; BRAATZ, 2000) como um desvio não permitido de ao menos um comportamento característico ou variável de processo. Aldrich e Auret (2013) definem falha como comportamento anômalo que causa sistemas ou processos desviarem inaceitavelmente das suas condições operacionais ou estados normais. A revisão provida por Venkatasubramanian et al. (2003) define falha como uma anormalidade ou sintoma, como temperatura elevada em um reator ou produto com baixa qualidade. Essa revisão também define a causa subjacente de uma falha, como uma bomba de refrigeração ou controlador com falha, como causa raiz ou evento básico. Como análise de causa raiz não é foco deste trabalho,

por simplicidade, todos esses termos são generalizados neste documento como anomalias ou eventos indesejados.

A tarefa de responder a anomalias em um processo envolve detecção e classificação em tempo hábil (prognóstico), diagnóstico de sua causa raiz e, então, implementação de ações apropriadas de controle para retornar o processo a um estado normal, seguro e operacional. Essa tarefa como um todo tem sido chamada de Gerenciamento de Eventos Anormais ou *Abnormal Event Management* (AEM) (VENKATASUBRAMANIAN *et al.*, 2003).

1.1.2 Sistemas de Elevação e Escoamento de Petróleo

Em todo lugar do mundo, tanto em ambientes marítimos quanto terrestres, extração de petróleo é feita através de estruturas denominadas poços de petróleo ou apenas poços (THOMAS, 2004).

Um poço de petróleo se refere a um conjunto de sensores e de sistemas mecânicos, elétricos e hidráulicos. Em poços marítimos, também chamados de poços *offshore*, essa estrutura pode ser parcial ou integralmente instalada no leito marinho, no fundo do poço e/ou na superfície (plataforma). Em poços terrestres, referenciados também como poços *onshore*, tais equipamentos são instalados na superfície da terra (DONNEZ, 2007) (THOMAS, 2004).

A principal função de um poço é interligar um reservatório de petróleo a uma instalação industrial. No caso de poços *offshore*, essa instalação industrial é uma unidade estacionária de produção (UEP), também referenciada como plataforma de produção de petróleo. Já no caso de poços *onshore*, essa instalação industrial é uma unidade de produção terrestre. Tanto as plataformas quanto as unidades de produção terrestre realizam apenas um pré-processamento do petróleo suficiente para separá-lo essencialmente entre óleo, gás e água enquadrados em especificações técnicas definidas pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). Em seguida, a água é reutilizada e/ou descartada e o óleo e o gás são transportados (exportados) para serem processados por, respectivamente, refinarias ou unidades de tratamento de gás (ANDREOLLI, 2016).

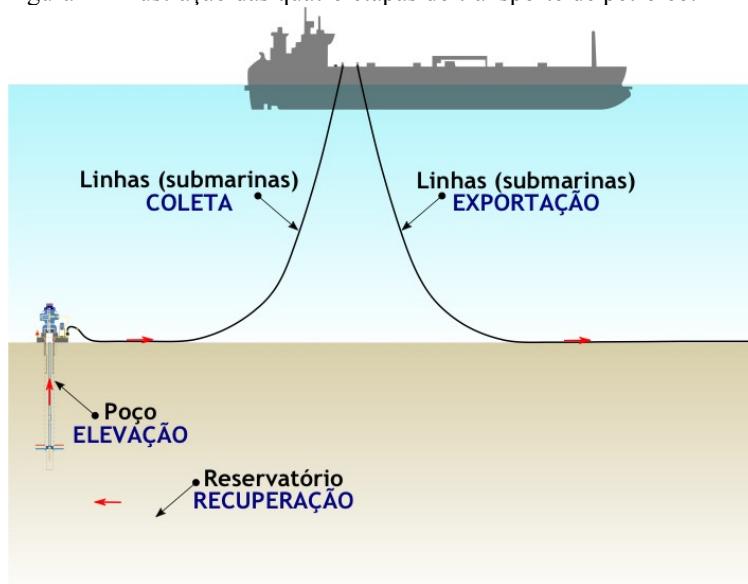
São dois os tipos primários de poços de petróleo: injetor e produtor. O primeiro é destinado a injeção de água, gás e/ou produtos químicos, a partir de uma instalação industrial, em um reservatório. A intenção desse tipo de operação é basicamente elevar a pressão no reservatório a fim de aumentar a produção dos seus poços. Apesar dessa contextualização, esta tese não considera esse tipo de poço. Já um poço produtor, como seu próprio nome revela, se destina a

produzir (transportar) petróleo desde um reservatório até uma instalação industrial (CAMPOS; GOMES; PEREZ, 2013) (ANDREOLLI, 2016). Neste documento, quando o tipo de poço não é especificado, trata-se de um produtor.

Elevação e Escoamento de Petróleo é a área de engenharia que reúne grande parte do conhecimento científico sobre o sub-processo, do processo mais amplo de extração de petróleo, dedicado a transportá-lo desde um reservatório até uma refinaria ou uma unidade de tratamento de gás (ANDREOLLI, 2016).

Esse tipo de transporte costuma ser dividido em quatro etapas, conforme apresentado na Figura 1. Essa divisão é particularmente interessante porque em cada etapa ocorrem diferentes fenômenos. Portanto, diferentes tipos de eventos considerados anomalias podem acontecer em cada uma dessas etapas (ANDREOLLI, 2016).

Figura 1 – Ilustração das quatro etapas do transporte de petróleo.



Fonte: (ANDREOLLI, 2016)

Nota: ilustração adaptada pelo autor.

As principais características de cada uma dessas etapas são as seguintes (ANDREOLLI, 2016):

1. Recuperação: o escoamento ocorre em meio poroso da rocha reservatório. Características da rocha e do fluido, composto basicamente por óleo, água, gás e sedimentos, são os fatores que mais influenciam nessa etapa do transporte;
2. Elevação: o escoamento ocorre na coluna de produção, que fica interna a um poço produtor. Como nessa etapa o escoamento se dá primordialmente na vertical, o que mais influencia no seu fluxo é a gravidade terrestre;

3. Coleta: essa etapa de escoamento é mais comum em poços marítimos. Ela ocorre entre um poço e uma plataforma através de tubulação que fica em contato direto com a água do mar. Em geral, toda essa tubulação é dividida em dois trechos, um praticamente horizontal (chamado de *flowline*) e outro praticamente vertical (chamado de *riser*). No primeiro, a fricção é o fator que atua de forma mais significativa no fluxo. No *riser*, as pressões costumam ser mais baixas, o que faz parte do gás se desprender do líquido e se expandir. Com isso, a fração de líquido do fluido diminui em relação às etapas anteriores e a componente gravitacional passa a influenciar menos no fluxo;
4. Exportação: o escoamento ocorre ou através de oleoduto ou através de gasoduto entre instalações industriais. Alguns dos possíveis casos são: entre uma plataforma e uma unidade de tratamento gás, entre uma plataforma e uma refinaria ou entre uma plataforma e um navio coletor de óleo e gás.

Em geral, a recuperação é especificada, simulada e acompanhada por profissionais de uma área da Engenharia de Petróleo chamada de Reservatório. Já as demais etapas, via de regra, ficam sob a responsabilidade de profissionais associados à área Elevação e Escoamento de Petróleo (ANDREOLLI, 2016).

Essa área define diferentes sistemas de elevação, que são conjuntos de técnicas de engenharia destinadas a elevação e escoamento de petróleo. Cada sistema é mais apropriado para determinadas conjunturas (pressão estática de reservatório, tipo de óleo, vazão desejada de produção, capacidade de pré-processamento e/ou escoamento, etc.). A quantidade e a diversidade de equipamentos necessários em cada sistema também variam.

Os sistemas de elevação e escoamento de petróleo mais comuns na atualidade, levando-se em consideração todo o globo, são brevemente descritos a seguir. Explicações mais abrangentes e detalhadas podem ser encontradas em (ANDREOLLI, 2016), (ROSA; CARVALHO; XAVIER, 2006) e (ROCHA; AZEVEDO, 2009).

Elevação Natural

Neste sistema, a pressão de formação, gerada pelo reservatório de petróleo, é suficiente para elevar e escoar hidrocarbonetos a uma taxa comercialmente aceitável sem requerer qualquer tipo de bombeamento. Ou seja, neste sistema não é requerida nenhuma fonte de energia além da contida no próprio reservatório na forma de pressão estática. Por conta disso, na indústria de

petróleo se diz que todo poço capaz de produzir (elevar e escoar) via elevação natural é sur gente. Essa terminologia também é utilizada em (THOMAS, 2004).

Uma característica marcante deste sistema em relação aos demais é o fato dele, em geral, demandar menos equipamentos. Consequentemente, a sua instrumentação (sensores e atuadores) também é, em geral, mais simples, o que resulta em uma operação com menor grau de liberdade e um monitoramento restrito a menos variáveis.

Elevação Artificial

Quando a pressão estática do reservatório não é suficiente para elevar e escoar o petróleo a uma taxa comercialmente aceitável, algum método artificial passa a ser necessário. São vários os sistemas de Elevação Artificial. Todos eles adicionam energia de alguma forma no reservatório e/ou no poço.

É importante destacar que poços capazes de produzir via elevação natural também podem ser equipados para serem operados em determinadas circunstâncias por métodos artificiais. Ou seja, não é incomum um poço ser operado ora por um método artificial ora pelo método natural.

Elevação Artificial com Injeção de Gás (*Gas Lift*)

A principal característica deste método é a injeção de gás natural em alta pressão, a partir da superfície, na coluna com fluido de produção através de uma ou mais válvulas submersas fixadas em profundidades pré-determinadas. Essa injeção pode ser contínua ou intermitente. Seu propósito fundamental é reduzir a massa específica média do fluido produzido para, assim, reduzir as perdas associadas à hidrostática do sistema de escoamento.

Elevação Artificial com Bombeio Centrífugo Submerso (BCS)

Neste método, uma bomba centrífuga que recebe energia de um motor elétrico provê energia adicional para que o fluido seja drenado. Essa classe de bomba pode ser instalada tanto na coluna de produção, o que configura o caso clássico, quanto no leito marinho, variante de BCS batizada de Bombeio Centrífugo Submerso Submarino (BCSS).

Elevação Artificial com Bombeio Mecânico (BM)

Este é o método mais utilizado no mundo. Contudo, sua aplicação é restrita a poços *onshore* com vazões de óleo de até 300 m³/d. Em resumo, este método transforma movimento rotativo

de um motor instalado na superfície em movimento alternativo em uma unidade de bombeio também instalada na superfície. Esse movimento alternativo é transmitido, então, por meio de hastes que deslizam internamente na coluna de produção, a uma bomba instalada no fundo do poço.

1.1.3 Estados de um Poço Produtor de Petróleo Operado via Elevação Natural

Durante a fase de produção de um determinado poço de petróleo, as equipes de monitoramento e os especialistas na área Elevação e Escoamento de Petróleo costumam dar nomes aos diferentes estados de operação sobre os quais ele pode se encontrar. Essa é uma prática ao menos na Petróleo Brasileiro S/A (Petrobras).

Alguns destes estados são associados a apenas certos métodos de elevação. Outros são associados a mais de um método.

Como ainda não há nenhuma norma, internacional ou nacional, ou livro de expressão que descreve tais estados, é esperado não haver consenso em relação aos termos utilizados e os seus significados específicos.

São cobertos nesta seção apenas os estados que são estritamente necessários a esta contextualização, associados à elevação natural e reconhecidos por praticamente todos os grupos de profissionais dessa área na Petrobras. Outros estados associados exclusivamente a outros métodos não são citados por não fazerem parte do escopo deste trabalho.

Os estados listados a seguir são mutuamente excludentes. Ou seja, cada poço pode se encontrar, necessariamente, em apenas um destes estados em um determinado instante. São eles:

- Fechado: nesse estado não há qualquer fluxo de petróleo através do poço produtor. Válvulas específicas são acionadas para fechá-lo. Isso é feito por motivos diversos, por exemplo: sinistro ou manutenção em plataforma ou na linha de produção;
- Operação normal: nesse estado, o poço se encontra em plena produção. Trata-se de um estado em regime e sem ocorrências significativas de anomalias;
- Partindo: esse é um estado transitório entre os dois estados já listados (fechado e operação normal). Diferentes conjunturas demandam diferentes procedimentos de partida de poço produtor;

- Fechando: esse é o estado transitório oposto ao partindo. Ou seja, um poço neste estado transita entre operação normal para fechado. Em linhas gerais, válvulas são fechadas até cessar completamente o fluxo produzido.

1.1.4 Aprendizado de Máquina e Ciência de Dados

Aprendizado de Máquina (*Machine Learning*) é um campo da Inteligência Artificial diretamente associada à Ciência da Computação, na qual também são utilizados conceitos e ferramentas de outras áreas do conhecimento, como Engenharia, Estatística, Probabilidade e Biologia.

A principal característica que torna um algoritmo pertencente a essa área é a existência de algum aprendizado sistematizado a partir de exemplos fornecidos e em geral consolidados nos chamados *datasets* (conjunto de dados) (FACELI *et al.*, 2011). A Seção 3.1 explica conceitualmente como esse tipo de aprendizado é obtido para as diferentes categorias de algoritmos de Aprendizado de Máquina, inclusive a de classificação, que é a mais relevante para este trabalho.

Por outro lado, a inteligência de um algoritmo convencional, que não realiza aprendizado sistematizado, precisa ser inserida de forma explícita pelos seus desenvolvedores, por exemplo no seu código-fonte ou em sua base de dados.

Uma expressão que tem se tornado cada vez mais popular é Ciência de Dados. Trata-se de um campo de pesquisa que busca abranger teorias e algoritmos de várias áreas que buscam extrair de dados algum tipo de conhecimento e com certo enfoque. Alguns exemplos são Análises de Dados (*Data Analysis*), Aprendizado Estatístico (*Statistical Learning*), Computação Cognitiva (*Computational Cognition*), Mineração de Dados (*Data Mining*) e Visualização de Dados (*Data Visualization*). Portanto, neste documento, teorias e algoritmos de Aprendizado de Máquina também são referenciados como de Ciência de Dados.

1.1.5 Ambiente de Desenvolvimento desta Tese

Este trabalho foi concebido e desenvolvido de forma aplicada e para beneficiar a indústria de petróleo em geral.

Ao longo do desenvolvimento desta tese, deu-se uma atenção especial à realidade da Petrobras, principalmente em relação às instalações da Unidade Operacional da Petrobras no Espírito

Santo (UO-ES). Importante frisar, contudo, que essa atenção especial não limitou a aplicabilidade dos resultados obtidos a essa companhia.

Faz-se relevante citar, inclusive, que este trabalho acadêmico impulsionou outras iniciativas isoladas e juntas deram origem na Petrobras a um projeto intitulado Monitoramento de Alarmes Especialistas (MAE). O seu foco de médio prazo é elevar o desempenho da Célula de Monitoramento e Vigilância de Poços do Centro de Operações Integradas (COI) da UO-ES. O seu efetivo planejamento se iniciou em janeiro de 2017 e em junho de 2017 ele foi formalizado junto à UO-ES. Ao final desse projeto, espera-se ter algoritmos para prognóstico de anomalias consideradas mais críticas em termos de potencial de impactar na produção de petróleo. Espera-se também que os seus resultados sejam reaproveitados em outras unidades da Petrobras e que outras rodadas desse projeto sejam iniciadas com foco em anomalias em outros processos da gerência executiva de Exploração e Produção de Petróleo (E&P) (falhas de equipamentos importantes de planta de processo, por exemplo). Assim, espera-se que Ciência de Dados e Aprendizado de Máquina sejam matérias cada vez mais exploradas em todas as áreas dessa empresa.

1.2 MOTIVAÇÃO

Esta seção introduz o problema de engenharia no qual esta tese se concentra. Um detalhamento desse problema é feito no Capítulo 2.

Como em praticamente qualquer processo industrial, há também no processo de elevação e escoamento de petróleo variados tipos de anomalias ou eventos indesejados. O problema central deste trabalho é justamente como detectar e classificar tais anomalias, por meio de monitoramento de variáveis de processo, em menor tempo e com maior desempenho.

As características gerais mais relevantes desse problema são as seguintes:

- Uma classificação identifica que ocorreu um tipo específico de anomalia;
- Uma detecção identifica que ocorreu alguma anomalia, sem especificá-la. Trata-se, portanto, de uma classificação binária¹ entre normalidade e anormalidade;
- Um prognóstico identifica antecipadamente um tipo específico de anomalia (antes do seu término). Em outras palavras, trata-se de classificação antecipada entre múltiplas classes;

¹ Em classificação binária, apenas dois rótulos são utilizados.

- Um diagnóstico identifica que ocorreu um tipo específico de anomalia e também as sua(s) causa(s). Trata-se, portanto, de classificação com indicação de sua(s) causa(s);
- Certas anomalias não ocorrem com frequência, mas em alguns casos apenas uma ocorrência é capaz de gerar efeitos negativos consideráveis;
- Certas anomalias só ocorrem em certos sistemas de elevação e escoamento de petróleo;
- Certas anomalias só ocorrem em poços marítimos e outras somente em poços terrestres;
- Mesmo com o processo em funcionamento, sensores monitorados podem se tornar indisponíveis.

Ao menos em todas as unidades de produção da Petrobras, no Brasil e nos outros países em que ela atua, esse problema vem sendo tratado fundamentalmente com a seguinte estratégia:

- Para detecção de anomalia, são desenvolvidos e atualizados múltiplos monitoramentos paramétricos e univariáveis, um para cada variável considerada relevante. Alarmes menos ou mais importantes são acionados a depender de quais parâmetros de monitoramento previamente configurados são violados. A detecção de anomalia é feita, portanto, por meio de múltiplos alarmes;
- Para classificação de anomalias, um profissional especialista na área Elevação e Escoamento de Petróleo analisa, após uma detecção de anomalia, se algum padrão característico já conhecido, formado pelas séries temporais das variáveis consideradas relevantes, se formou ou está em formação. A classificação de anomalias é realizada, então, por conhecimento especialista e cognição humana.

1.3 HIPÓTESES

São duas as hipóteses formuladas para esta tese. A saber:

1. Anomalias em poços de petróleo podem ser detectadas com Aprendizado de Máquina;
2. Treinar algoritmos de Aprendizado de Máquina com exemplos de anomalias gerados por simulador computacional ou desenhados à mão por especialistas é uma solução viável para detecção de ocorrências reais raras² dessas anomalias.

Ambas as hipóteses se restringem a anomalias que ocorrem prioritariamente em poços produtores marítimos operados com Elevação Natural, que se iniciam durante operação normal

² O limiar utilizado para diferenciar anomalias raras foi de 1%, o mesmo critério utilizado em (ZHANG; ZHOU, 2007). Com esse critério, todo tipo de anomalia cujo número de ocorrências reais representa menos de 1% dentre todas as ocorrências reais incluídas no *dataset*, inclusive as de normalidade, é considerado raro.

e que se evidenciam nas etapas de recuperação, elevação e/ou coleta. Anomalias que ocorrem exclusivamente em outros sistemas de elevação e escoamento de petróleo não são consideradas nessas hipóteses.

1.4 OBJETIVOS

O objetivo geral desta tese é evidenciar que as hipóteses formuladas na Seção 1.3 são verdadeiras. Para isso, foi planejado e desenvolvido um trabalho com escopo restrito aos seguintes objetivos específicos:

- Estudar o problema descrito no Capítulo 2 no âmbito da Petrobras;
- Revisar a fundamentação teórica coberta no Capítulo 3;
- Conceber e aplicar no contexto da Petrobras a metodologia detalhada no Capítulo 4 e destinada à (i) preparação de um *dataset* composto por ocorrências de anomalias reais históricas, simuladas e desenhadas à mão por especialistas na área Elevação e Escoamento de Petróleo e à (ii) proposição de desafios específicos (*benchmarks*) que testam as hipóteses estabelecidas neste trabalho;
- Experimentar algoritmos disponíveis publicamente nos *benchmarks* propostos e obter resultados que sugiram que as tais hipóteses são verdadeiras, conforme apresentado no Capítulo 5.

1.5 JUSTIFICATIVAS

Nesta seção, são apresentadas a relevância do problema introduzido na Seção 1.2, o que legitima a sua escolha neste trabalho, e as justificativas tanto para as formulações das hipóteses apresentadas na Seção 1.3 quanto para os objetivos específicos, listadas na Seção 1.4, que definem o escopo desta empreitada.

Detecção, classificação e prognóstico de anomalias raras são tarefas relevantes e em voga em diversas atividades realizadas e/ou monitoradas pelos humanos. Alguns exemplos são detecção de influxo durante perfuração (TANG *et al.*, 2019); detecção e localização de vazamento em dutos de água ou oleodutos (LIU; LI; XU, 2019); detecção de falhas em plantas industriais (ARRUDA *et al.*, 2014) (HE; WANG, 2007) (XAVIER; SEIXAS, 2018), em poços de petróleo (LIU *et al.*, 2011) (LIU *et al.*, 2010) (LIU *et al.*, 2010), em BCS (PATRI *et al.*, 2014) e em válvulas de compressores de gás (PATRI *et al.*, 2015) (PATRI *et al.*, 2016); detecção de eventos anormais em poços de petróleo (SANTOS *et al.*, 2018) (VARGAS *et al.*, 2017), em

Eletroencefalograma (EEG) (UNG *et al.*, 2017) e em Eletrocardiograma (ECG) (BAGNALL *et al.*, 2017); detecção de distúrbios em qualidade de energia (GRAY; MORSI, 2015) (GAING, 2004); e detecção de armas de mão (OLMOS; TABIK; HERRERA, 2018).

Prognóstico e diagnóstico em um GEA automatizado podem ser vistos como problemas de classificação e algoritmos de classificação podem ser categorizados em termos dos seus conhecimentos (modelos) e estratégias de busca (VENKATASUBRAMANIAN *et al.*, 2003).

A aplicação de algoritmos de Aprendizado de Máquina a dados obtidos de processos é uma estratégia de busca que tem sido utilizada com sucesso, especialmente nos últimos anos. Métodos de Aprendizado de Máquina têm sido aplicados em diferentes tarefas, como predição de condições de trabalho em fundo do poço (LI; HAN; WANG, 2018), classificação de modelos bem testados a partir de dados de testes de transiente de pressão (AHMADI; AMINSHAHIDY; SHAHRABI, 2017), análise automática de dados de perfuração em tempo real e detecção de eventos relacionados a influxo (TANG *et al.*, 2019). Em (BHATTACHARYA; MISHRA, 2018), algoritmos de Aprendizado de Máquina são aplicados para classificação de fácies e fraturas em reservatórios convencionais e não convencionais. Em (XAVIER; SEIXAS, 2018), é proposto um método que permite diagnóstico de falhas, assim como detecção, sem parâmetros ajustáveis e que pode ser usado com grandes volumes de dados com pouca informação útil. Métodos para detecção de falhas são propostos em (ARRUDA *et al.*, 2014) (HE; WANG, 2007). Um método que detecta acumulação de hidrato em linhas de produção e injeção de poços de petróleo com no mínimo uma hora de antecedência é apresentado em (SANTOS *et al.*, 2018). Uma transformação específica em cada variável monitorada é realizada em (VARGAS *et al.*, 2017) antes da classificação em si por um algoritmo de Aprendizado de Máquina para evitar recalibração de modelo mesmo em sistemas dinâmicos. Classificação de séries temporais é empiricamente avaliada em 39 *datasets* em (LI *et al.*, 2016). Métodos para predição de falha em poços de petróleo são propostos em (LIU *et al.*, 2011) (LIU *et al.*, 2010) (LIU *et al.*, 2010). Outros trabalhos relacionados a algoritmos para classificação de séries temporais são publicados em (PATRI *et al.*, 2014) (PATRI *et al.*, 2015) (GEURTS, 2001) (XI *et al.*, 2006).

Um efetivo GEA em poços de petróleo ajuda a reduzir custos de manutenção e a prevenir perdas de produção³, assim como acidentes ambientais e à vida. Essa atividade se torna ainda mais importante em poços marítimos devido suas produções, riscos envolvidos e custos serem

³ Perda de produção de petróleo é o termo utilizado quando se deixa de extrair esse produto durante certo período.

geralmente maiores (ANDREOLLI, 2016) (CAMARGO *et al.*, 2010). O catastrófico acidente Macondo, que ocorreu em 2010 devido a uma falha em um equipamento de segurança, exemplifica a magnitude potencial de perdas e custos nesse contexto. A morte de onze trabalhadores, o afundamento da sonda *Deepwater Horizon* e os danos marinhos e costeiros marcaram esse incidente como um dos maiores desastres ambientais na história dos Estados Unidos da América (SUTHERLAND *et al.*, 2016). Em termos de manutenção, o custo de uma sonda marítima para reparar uma linha de produção, por exemplo, pode exceder US\$500 mil por dia (ANDREOLLI, 2016).

Por conta da natureza mutável do sistema de produção de petróleo, especialmente em relação ao reservatório de produção e aos fenômenos decorrentes do fluxo multifásico, o clássico monitoramento paramétrico e univariável, como parte de um GEA, se mostra oneroso e menos eficiente em comparação com uma abordagem multivariável (HARIDY; WU, 2009). Nesse tipo de monitoramento, se as várias e complexas reparametrizações necessárias ao longo da vida útil de um poço não são feitas, sua eficiência é comprometida.

Outras estratégias de monitoramento usam modelos caixa-branca, que são construídos a partir de conhecimento dos fenômenos do sistema e dos modelos (exemplos: físico-químico, termodinâmico e fluxo multifásico) (PIANTANIDA *et al.*, 2013). Suas acurácia dependem substancialmente da aderência do modelo e da qualidade da identificação dos seus parâmetros. Devido ao custo de aquisição de licenças de simuladores e dos seus tempos de processamento, geralmente essa abordagem não é largamente utilizada na indústria para monitoramento *online*.

A Petrobras também considera a relevância de GEA em poços de petróleo, que é encarado como um catalizador eficiente que aumenta produção e reduz diversos riscos e custos. A UO-ES tem priorizado há alguns anos investimentos para substituir seus monitoramentos paramétricos por não paramétricos. Esse processo começou em 2015, exclusivamente com monitoramentos de poços marítimos operados por Elevação Natural. Classificadores baseados em Redes Neurais Artificiais foram projetados, desenvolvidos e experimentados principalmente pela então existente gerência UO-ES/ENGP/EES. Os resultados práticos obtidos foram considerados bons na época, mas percebeu-se que eles poderiam ser melhores se fossem utilizadas metodologias voltadas para preparação de base de dados e para desenvolvimento, seleção e avaliação de algoritmos. Essa constatação influenciou decisivamente na escolha do tema desta tese. Mais recentemente, a UO-ES iniciou com o seu projeto MAE uma evolução do seu GEA dedicado a poços de petróleo. Isso em conjunto com esta pesquisa e influenciada por ela. O foco

estabelecido nessa evolução foi evoluir o seu GEA, atualmente baseado em alarmes univariáveis paramétricos, com a utilização de algoritmos de Aprendizado de Máquina aplicados a Série Temporal Multivariável (STM), ou *Multivariate Time Series* (MTS), obtidas do processo industrial. A definição formal de MTS se encontra na Seção 3.2.

A consideração da estratégia de busca baseada em aplicação de algoritmos de Aprendizado de Máquina na primeira hipótese estabelecida neste trabalho é justificada principalmente pelo fato dessa estratégia estar sendo utilizada com sucesso nos últimos anos. Outra justificativa para essa consideração é a alta disponibilidade de MTS relacionadas a variáveis de processo. Contudo, mesmo após uma extensa pesquisa, nenhum *dataset* público ou privado com quantidade e diversidade suficientes de anomalias em poços de petróleo foi encontrado.

Um dos *datasets* privados encontrados é parte do *Electric Submersible Pump – Reliability Information and Failure Tracking System* (ESP-RIFTS) (C-FER TECHNOLOGIES, 2019), desenvolvido pelo *Joint Industry Project* (JIP) e mantido pela C-FER Technologies (C-FER TECHNOLOGIES, 2019). O objetivo do ESP-RIFTS é incrementar significativamente a vida útil de BCS. O seu conjunto de dados é composto por informações extraídas de aproximadamente 112 mil BCS instaladas em torno de 800 campos operados por 26 companhias ao redor do mundo. Contudo, esse conjunto de dados não contém nenhuma série temporal associada a variáveis de processo. Apenas dois *datasets* com MTS obtidas a partir de processo foram encontrados, ambos de propriedade da Petrobras e privados. O primeiro é composto por 11 ocorrências (quatro simuladas e sete reais históricas) de quatro tipos de anomalias e foi explorado em (VARGAS *et al.*, 2017). O segundo contém 12 ocorrências reais históricas associadas a apenas um tipo de anomalia e foi utilizado em (SANTOS *et al.*, 2018). Mesmo combinados, entendeu-se que esses dois *datasets* não têm nem quantidade nem diversidade suficiente de anomalias.

Em repositórios públicos (DHEERU; CASEY, 2019) (DAU *et al.*, 2018) (BAGNALL *et al.*, 2019) (CHEN, 2019) (ISTITUTO NAZIONALE DI STATISTICA, 2019), apenas *datasets* associados a outros contextos foram encontrados. Além disso, em geral eles são pré-tratados e se tornam irrealistas em algum nível. Alguns exemplos de pré-processamento são eliminação ou substituição de valores ausentes (*missing*) ou de variáveis congeladas (cujos valores se mantêm fixos por conta de problema em sensor, configuração de sistema ou em rede de comunicação); consideração de séries temporais com apenas tamanhos iguais ou apenas dados simulados; e balanceamento em relação a quantidade de ocorrências por tipo de anomalia.

O resultado dessa investigação ratificou a necessidade de uma metodologia ser concebida e aplicada para a preparação de um conjunto de dados próprio e realista. Essa é a origem do 3W dataset⁴, até onde os seus autores têm conhecimento, o primeiro *dataset* público com dados de anomalias reais em poços de petróleo.

Outras justificativas aplicáveis tanto a este trabalho quanto ao projeto MAE que merecem ser explicitadas são as seguintes:

- Restringiu-se a anomalias que ocorrem prioritariamente em sistema de elevação natural por ele ser o mais clássico e menos complexo e por conta de parte dessas anomalias ocorrerem também em outros sistemas de elevação e escoamento de petróleo;
- Focou-se em prognóstico porque ações de tratamento mais rápidas e adequadas em ocorrências dessas anomalias, em geral, mitigam de forma mais eficiente perdas de produção, custos e acidentes;
- Para certos tipos de anomalias considerados neste trabalho, por exemplo incrustação em CKP (*Choke de Produção*) e hidrato em linha de serviço, a identificação da sua causa é direta e revelada pelo seu próprio nome. Contudo, esse tipo de indicação para outros tipos de anomalias, por exemplo perda rápida de produtividade e instabilidade de fluxo, demanda métodos específicos. Por conta disso, diagnóstico de anomalias não faz parte do escopo nem do projeto MAE nem desta tese;
- Ocorrências de anomalias simuladas e desenhadas à mão por especialistas na área Elevação e Escoamento de Petróleo foram incluídas no 3W dataset como uma solução viável para enriquecimento do seu conhecimento *a priori*;
- Utilizou-se o PI System (OSISOF, 2019) por ele ser o *plant information system* utilizado na UO-ES;
- Utilizou-se o simulador OLGA (SCHLUMBERGER, 2019) por ele ser um dos poucos sistemas que simulam fenômenos transientes associados a extração de petróleo (ANDREOLLI, 2016) (GRØDAHL, 2014) e por ele ser a ferramenta padrão na Petrobras para simulação de situações em poços de petróleo;
- Utilizou-se uma ferramenta desenvolvida no âmbito do projeto MAE para digitalização de ocorrências desenhadas à mão porque nenhuma outra adequada para esse propósito foi encontrada;

⁴ O nome 3W dataset foi escolhido por ele ser composto por instâncias de 3 origens diferentes e que contêm anomalias que ocorrem em poços de petróleo (*Wells*).

- Ambas as hipóteses se restringem a anomalias cujas ocorrências se iniciam durante operação normal porque se espera que os poços se encontrem majoritariamente nesse estado e também por uma questão de delimitação de escopo para este trabalho;
- A segunda hipótese foca em anomalias raras porque para classificação de anomalias reais frequentes, ocorrências reais dessas anomalias podem ser divididas satisfatoriamente em subconjuntos para treinamentos, validações e testes;
- Foram experimentados apenas algoritmos disponíveis publicamente nos *benchmarks* propostos para que os seus resultados sirvam de base para comparações com resultados de algoritmos especializados que venham a ser desenvolvidos.

Como parte do projeto MAE, um levantamento na UO-ES estimou em 1.514.000 bbl (barris) a perda de produção acumulada no decorrer de 2016 causada por anomalias em seus poços produtores marítimos operados por Elevação Natural. Se essa estimativa é considerada como média anual, a US\$ 50/bbl, o impacto financeiro dessas anomalias na UO-ES é de US\$ 75,7 milhões. Esse levantamento não incluiu, contudo, outros resultados negativos decorrentes dessas anomalias que também justificam os investimentos no projeto MAE e a realização deste trabalho de pesquisa, tais como:

- Manutenção emergencial em equipamentos de poços, que às vezes demanda utilização de sonda marítima, um recurso considerado escasso e caro;
- Manutenção emergencial em equipamentos importantes de instalações industriais (em plataformas, por exemplo);
- Envolvimento de equipes e especialistas que deixam de atuar em outras frentes para tratar consequências dessas anomalias;
- Ocorrência de acidentes, com possíveis multas e impactos à vida ou ao meio ambiente.

1.6 CONTRIBUIÇÕES

Este trabalho deu origem a duas contribuições relevantes.

A primeira é o 3W dataset em si, que se encontra disponível publicamente em (VARGAS *et al.*, 2019, 2019a) e pode ser utilizado como um conjunto de dados de referência para desenvolvimento de técnicas de Aprendizado de Máquina para tratamento de dificuldades inerentes a dados reais. Um grande número de possibilidades em termos de, por exemplo, pré-processamento (normalização, valores ausentes, variáveis congeladas, *outliers*, etc.), filtros (suavização, reamostragem, etc.), transformações (escala múltipla temporal, *wavelet*, etc.),

família de classificadores (baseados em árvores, redes neurais artificiais, distâncias, etc.), otimização de hiperparâmetros, engenharia de características e métricas de desempenho podem ser investigadas utilizando-se esse *dataset*. Além disso, essas possibilidades podem ser exploradas em tarefas específicas associadas a classificação de anomalias em poços de petróleo e gás marítimos surgentes, como detecção de novas falhas (KRAWCZYK *et al.*, 2017); e classificação antecipada (HE *et al.*, 2013) (XING *et al.*, 2011) (XING; PEI; YU, 2009), de classe única (KRAWCZYK *et al.*, 2017), binária (KRAWCZYK *et al.*, 2017), múltiplas classes (KRAWCZYK *et al.*, 2017), *online* e *offline*⁵. Mais informações sobre cada uma dessas tarefas podem ser obtidas em (ALDRICH; AURET, 2013) (JAMES *et al.*, 2013) (WITTEN; FRANK; HALL, 2011) (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) (BISHOP, 2006) (DUDA; HART; STORK, 2001).

A segunda contribuição desta pesquisa é a especificação de desafios (*benchmarks*) que praticantes e pesquisadores podem utilizar em conjunto com o 3W dataset.

1.7 ORGANIZAÇÃO DA TESE

O restante deste documento é organizado da seguinte forma. O problema de engenharia que motivou esta pesquisa é caracterizado no próximo capítulo. O Capítulo 3 concentra os principais fundamentos teóricos utilizados neste trabalho. A metodologia desenvolvida e aplicada na preparação do 3W dataset e na proposição dos *benchmarks* é detalhada no Capítulo 4. O Capítulo 5 relata os resultados obtidos. No Capítulo 6, são sintetizadas as conclusões desta tese e sugeridos trabalhos futuros.

⁵ No contexto deste trabalho, classificação *offline* é a tarefa cujo objetivo é estimar uma única classe para a MTS inteira. Um aprendizado em lote que extrai características apenas de MTS inteiras é utilizado. Por outro lado, em classificação *online*, predição de classe e extração de características são feitas em múltiplas subpartes de cada MTS.

2 DESCRIÇÃO DO PROBLEMA

Este capítulo descreve o problema introduzido na Seção 1.2. Descrição essa que se limita ao contexto considerado nos objetivos traçados para esta pesquisa. Portanto, são apresentados aqui os poços marítimos surgentes de petróleo, os sensores geralmente disponíveis nesse sistema, os tipos de anomalias mais críticas que ocorrem prioritariamente nesses poços e os tamanhos de janelas temporais utilizados por especialistas para confirmar ocorrências de anomalias.

2.1 POÇOS MARÍTMOS SURGENTES DE PETRÓLEO

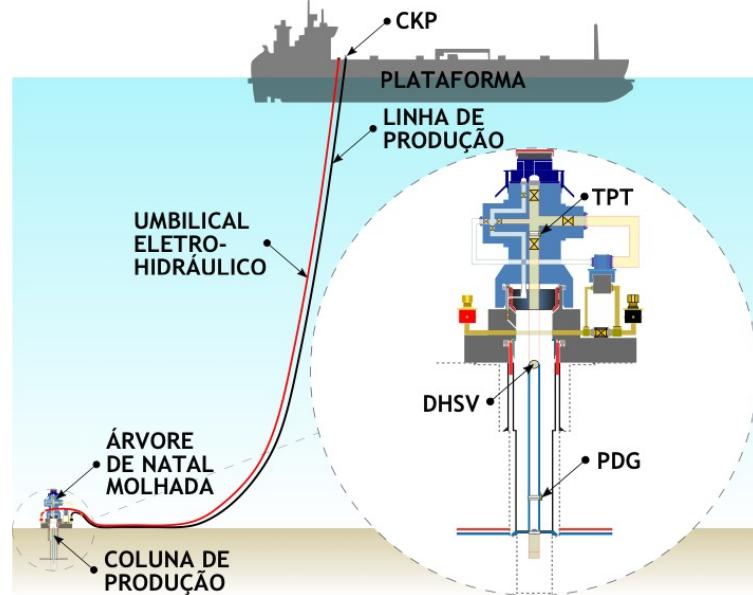
A Figura 2 traz um esquema simplificado de um poço marítimo típico operado com esse sistema. Estão presentes nessa ilustração apenas os componentes necessários a esta descrição. Mais informações podem ser encontradas em (ANDREOLLI, 2016), (CAMPOS; GOMES; PEREZ, 2013) e (THOMAS, 2004). Sobre esses componentes, é válido comentar que:

- O petróleo flui a partir do reservatório pela Coluna de Produção, Árvore de Natal Molhada (ANM), Linha de Produção e pela válvula *Choke* de Produção (CKP), nessa ordem, até a plataforma;
- ANM é um equipamento composto por basicamente válvulas operadas remotamente e sensores, inclusive o *Temperature and Pressure Transducer* (TPT);
- CKP é uma válvula de controle instalada no início da unidade de produção e é responsável pelo controle do poço na superfície;
- *Permanent Downhole Gauge* (PDG) e TPT são equipamentos que contêm sensores de pressão e de temperatura;
- *Down Hole Safety Valve* (DHSV) é um tipo de válvula hidráulica de segurança do tipo falha-fecha;
- O Umbilical Eletro-Hidráulico interliga as válvulas e sensores da ANM, assim como o PDG e a DHSV, aos sistemas na superfície.

Cada poço produtor é monitorado por meio de múltiplos sensores de variáveis de processo, principalmente de pressão e de temperatura. Sucessivas leituras desses sensores dão origem a MTS que ajudam a descrever o comportamento de cada poço.

O problema em si que motivou este trabalho é como detectar e classificar certas anomalias em poços produtores marítimos operados por Elevação Natural, por meio de monitoramento dessas MTS, em menor tempo e com maior desempenho.

Figura 2 – Esquema de um poço marítimo surgente de petróleo.



Fonte: próprio autor.

2.2 SENSORES GERALMENTE DISPONÍVEIS

Os sistemas de elevação e escoamento de petróleo possuem uma quantidade pequena de sensores em comparação às instalações de superfície devido aos altos custos e às dificuldades de acesso. Por exemplo, alguns sensores são instalados em profundidades de lâmina d'água maiores que mil metros e no interior da coluna de produção.

Esses sensores nem sempre estão bem calibrados e em certas situações eles permanecem em falha por longos períodos até serem substituídos ou calibrados. Em alguns casos, não há rotina de calibração por conta do alto custo de intervenção.

Os tipos e a quantidade de sensores utilizados em cada sistema de elevação e escoamento de petróleo podem variar dependendo da sua configuração. No caso típico de poço satélite de completação molhada operando por Elevação Natural, geralmente estão disponíveis até cinco sensores relevantes contidos em três equipamentos diferentes: PDG, TPT e CKP. As posições desses equipamentos no circuito de produção são apresentadas na Figura 2.

Esses sensores e suas principais características são descritos a seguir:

- Pressão do fluido no PDG (P-PDG): sensor mais próximo do reservatório e que usualmente falha devido ao nível de hostilidade do ambiente. Além disso, como o PDG é um equipamento enroscado na própria coluna de produção, em caso de falha desse

sensor, a sua substituição é considerada custosa e arriscada, pois demanda a substituição da própria coluna de produção;

- Pressão e temperatura do fluido no TPT (P-TPT e T-TPT): sensores internos à ANM que têm confiabilidade considerada boa pelos profissionais da área Elevação e Escoamento de Petróleo;
- Pressão do fluido montante à válvula *Choke* de Produção (P-MON-CKP): sensor também considerado de boa confiabilidade, quando presentes;
- Temperatura do fluido jusante à válvula *Choke* de Produção (T-JUS-CKP): nessa posição podem existir fluidos de vários poços diferentes, mas esse sensor é considerado relevante porque em geral não há sensor de temperatura do fluido montante à *Choke* de Produção.

Deve-se notar que, apesar de todos esses sensores terem sua importância, os de pressão no fluido no PDG e no TPT costumam ser os mais relevantes para as análises em relação a elevação e escoamento de petróleo. E são justamente esses equipamentos, em especial o PDG, que costumam falhar ou até mesmo faltar em certos poços.

2.3 ANOMALIAS MAIS CRÍTICAS

Conforme explicado na Subseção 1.1.2, transporte de petróleo costuma ser dividido em quatro etapas: recuperação, elevação, coleta e exportação. Nesta seção, são brevemente descritas apenas parte dos tipos de anomalias mais críticas (em termos de potencial de impactar na produção de petróleo) que podem ocorrer nessas três primeiras etapas. Descrições mais detalhadas podem ser encontradas em (ANDREOLLI, 2016), (SCHMIDT; DOTY; DUTTA-ROY, 1985), (ABASS; BASS, 1988) ou (ELLISON; GALLAGHER; LORIMER, 2000).

Em função do escopo desta tese, as subseções a seguir cobrem apenas anomalias que ocorrem prioritariamente em poços produtores marítimos operados via Elevação Natural.

A descrição aqui feita cobre quatro propriedades de cada tipo de anomalia:

- Fenômenos que explicam as dinâmicas típicas das séries temporais referentes às variáveis monitoradas;
- Consequências negativas mais frequentes;
- Reversibilidade. Ou seja, se há ações viáveis para reverter a anomalia com objetivo de mitigar perdas de produção, custos e riscos à vida e ao meio ambiente.

Para ilustrar essas propriedades, em cada subseção a seguir é apresentada uma figura com as séries temporais associadas aos sensores geralmente disponíveis durante uma ocorrência real da anomalia descrita. Todas essas ocorrências de anomalias fazem parte do 3W dataset. Para sensores indisponíveis à época, a série temporal é omitida. Conforme detalhado na Seção 4.1, rótulos no nível de observações foram utilizados para a determinação dos períodos indicados na legenda de cada figura.

2.3.1 Aumento Abrupto de BSW

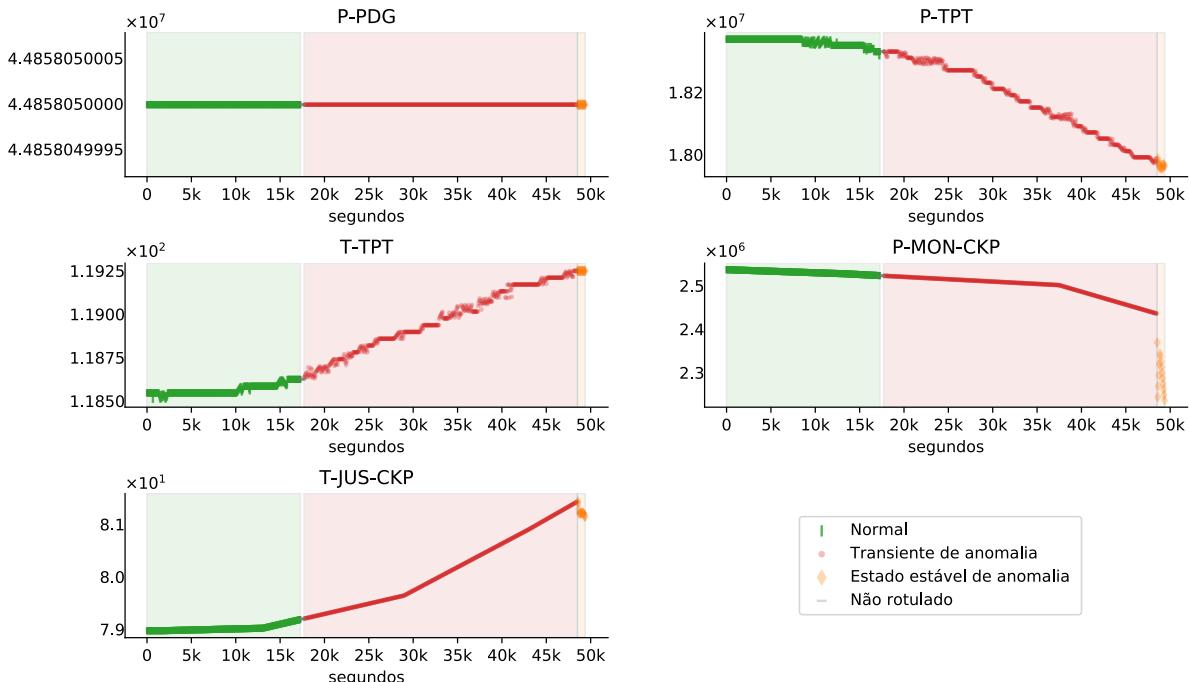
O *Basic Sediment and Water* (BSW) é definido como a razão entre a vazão de água mais sedimentos produzidos e a vazão de líquido produzido, ambas medidas em condições padrão (ANDREOLLI, 2016) (ABASS; BASS, 1988).

Projetos de tubulações de poços levam em consideração informações de BSW, que são inicialmente obtidas por modelagem de reservatório. Durante a vida útil de um poço, é esperado que o seu BSW aumente, por conta da maior produção de água, proveniente seja do aquífero natural do reservatório ou da injeção artificial para evitar declínio da sua produção. Contudo, um aumento abrupto de BSW pode acarretar diversos problemas relacionados a garantia de escoamento, a produção menor de óleo, a elevação do petróleo, ao processamento na instalação industrial e ao fator de recuperação (ANDREOLLI, 2016).

Mudanças no BSW podem afetar diversas variáveis em um sistema de elevação e escoamento de petróleo, com efeitos positivos e/ou negativos para a produtividade do poço. Por causa disso, as pressões e as temperaturas medidas podem tanto aumentar, quanto diminuir. A regra geral, entretanto, é que as pressões medidas em pontos mais profundos aumentem e que as pressões nos pontos mais próximos à superfície diminuam. As temperaturas, no geral, costumam aumentar em todos os equipamentos.

A Figura 3 ilustra uma das possíveis dinâmicas nas variáveis monitoradas quando da ocorrência deste tipo de anormalidade. É importante observar que nesse caso as variáveis entraram em estado estável de anomalia após cerca de 8 horas e 30 minutos desde o início da anormalidade indicado por um especialista da área. Uma reflexão sobre a constância da leitura de pressão no PDG se faz necessária. Como a medição ficou congelada em valor diferente de zero, há ao menos duas possíveis explicações para essa constância: sensor danificado ou valor inadequado de exceção configurado na fonte desses dados (supervisório) e/ou no PI System.

Figura 3 – Gráficos relacionados a um exemplo de ocorrência real de aumento abrupto de BSW. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFIT, 2019).

Nota: gráfico gerado pelo autor.

2.3.2 Fechamento Espúrio de DHSV

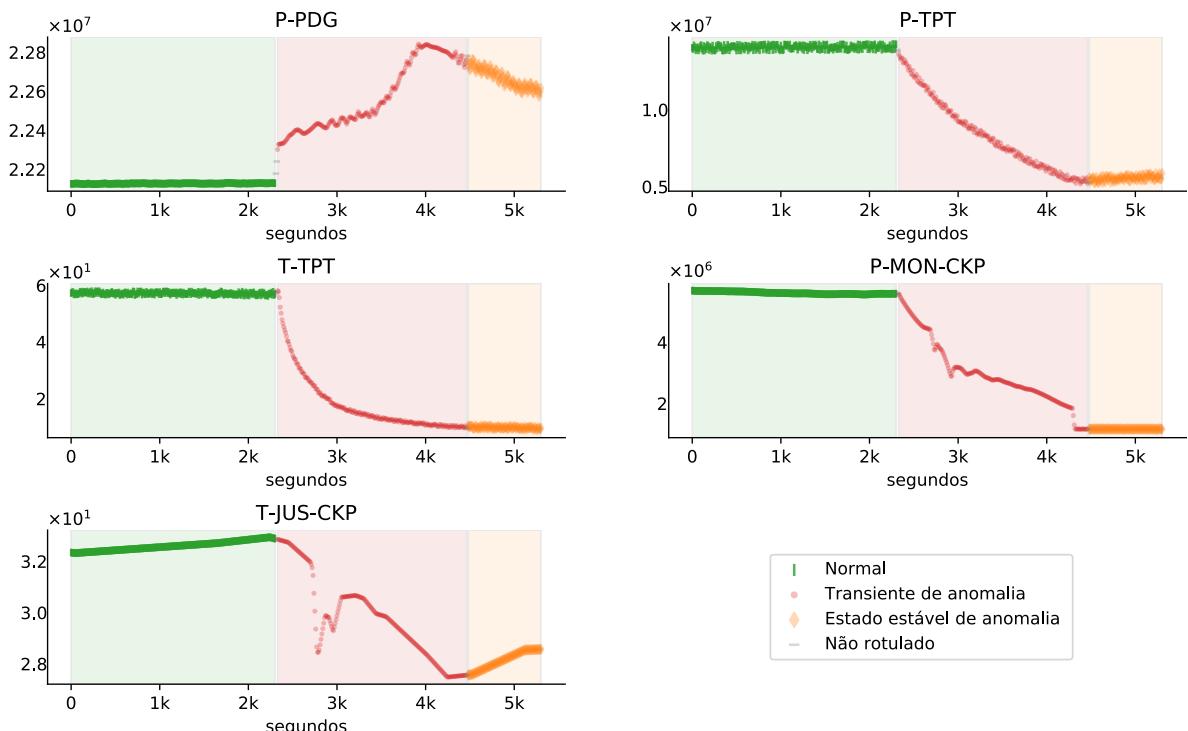
DHSV, também referenciada por apenas *Downhole Safety Valve* (DSV), é uma válvula de segurança que se instala na coluna de produção de um poço. Seu objetivo é assegurar o fechamento do poço em caso de sinistro em que a unidade de produção e o poço são fisicamente desconectados. Essa válvula é mantida aberta por um atuador hidráulico e se fecha em caso de desconexão entre o poço e a unidade de produção. (SCHLUMBERGER, 2019) (STANDARDS NORWAY, 2013).

Eventualmente, a DHSV falha e fecha de forma espúria, em geral sem qualquer indicação na superfície (queda de pressão no atuador hidráulico, por exemplo).

Uma identificação automática de fechamento espúrio dessa válvula em tempo hábil (prognóstico) permite sua reabertura por meio de procedimentos operacionais corretivos, o que evita perda de produção e custos adicionais.

Um caso real considerado representativo desta anomalia é apresentado na Figura 4. É importante observar que nesse caso o poço levou cerca de 35 minutos para entrar estado estável de anomalia (parada total de produção).

Figura 4 – Gráficos relacionados a um exemplo de ocorrência real de fechamento espúrio de DHSV. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFIT, 2019).

Nota: gráfico gerado pelo autor.

2.3.3 Intermitência Severa

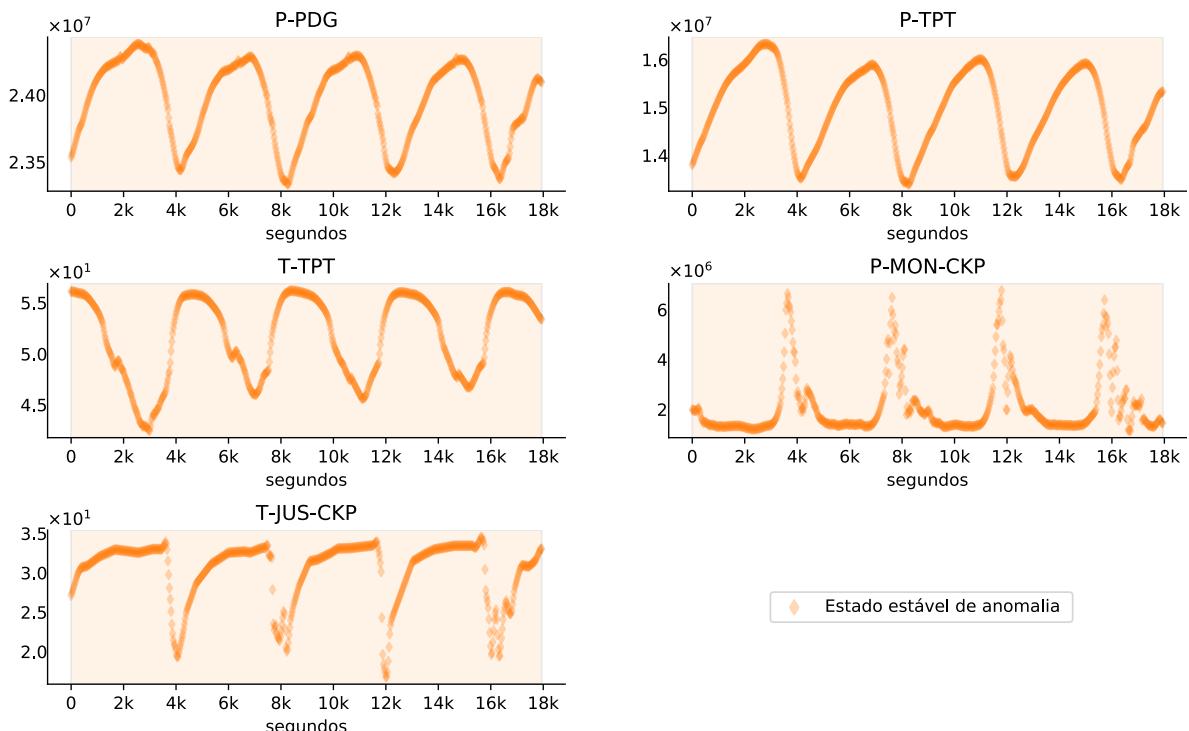
Trata-se de uma instabilidade mais crítica. Suas características mais marcantes são (MEGLIO *et al.*, 2012) (SCHMIDT; DOTY; DUTTA-ROY, 1985):

- Periodicidade bem definida;
- Intensidade suficiente para, em geral, poder ser detectada por sensores ao longo de todo o circuito de produção.

A depender da periodicidade e da intensidade, esse tipo de anomalia pode resultar em estresse ou mesmo danos em equipamentos do próprio poço e/ou da planta industrial. Contudo, um prognóstico desse tipo de anomalia permite ações específicas na operação do poço destinadas a reverter o quadro.

A Figura 5 traz um caso real deste tipo de anomalia. Como 18.000 segundos compreendem cerca de 4,5 períodos de intermitência, pode-se estimar que a periodicidade neste caso foi de, aproximadamente, 67 minutos.

Figura 5 – Gráficos relacionados a um exemplo de ocorrência real de intermitência severa. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFT, 2019).

Nota: gráfico gerado pelo autor.

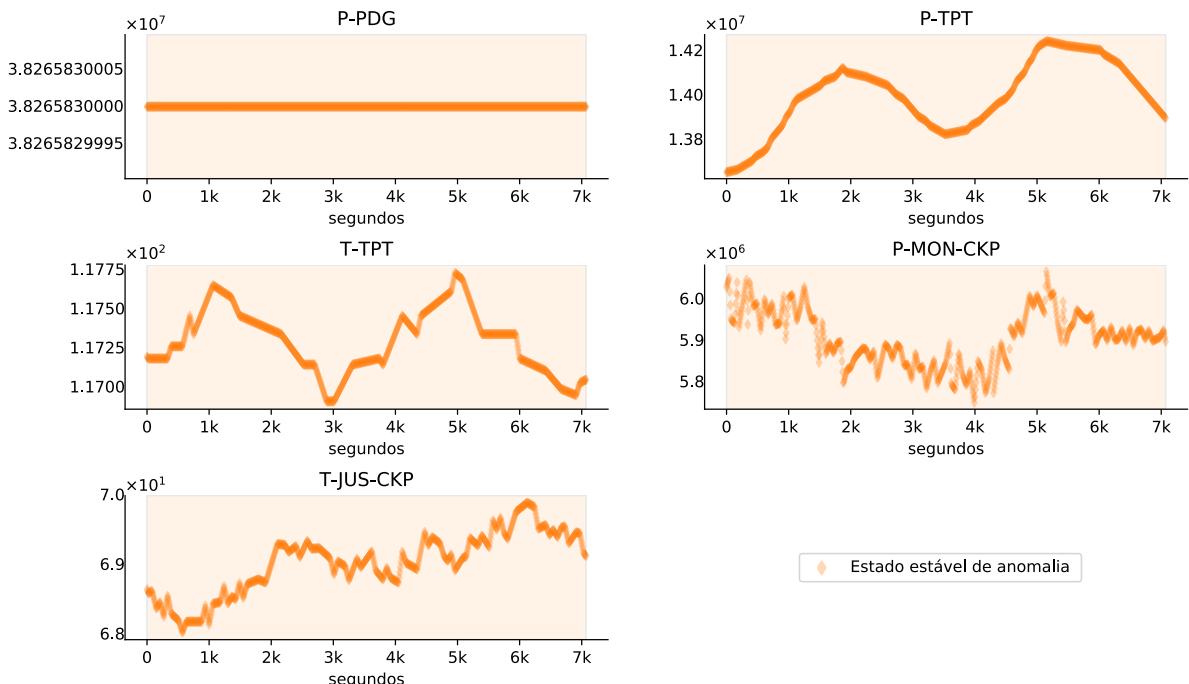
2.3.4 Instabilidade no Fluxo

Durante uma instabilidade no fluxo, ao menos parte das séries temporais monitoradas apresentam alterações de amplitude relevantes, mas toleráveis. Uma característica que diferencia este tipo de anomalia em relação a intermitência severa é a falta de periodicidade entre essas alterações (THEYAB, 2018) (TAKEI *et al.*, 2010).

Como uma instabilidade pode evoluir para uma intermitência severa, o seu prognóstico evita todos os aspectos negativos associados a essa anomalia mais grave.

Um caso real de instabilidade é apresentado na Figura 6, no qual inclusive o P-PDG se encontrava congelado.

Figura 6 – Gráficos relacionados a um exemplo de ocorrência real de instabilidade no fluxo. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFT, 2019).

Nota: gráfico gerado pelo autor.

2.3.5 Perda Rápida de Produtividade

A produtividade em um poço surgente depende de uma série de propriedades relativas ao reservatório: pressão estática, percentual de sedimentos e água, índice de produtividade, razão gás/óleo, viscosidade do fluido produzido, etc. (HAUSLER; KRISHNAMURTHY; SHERAR, 2015).

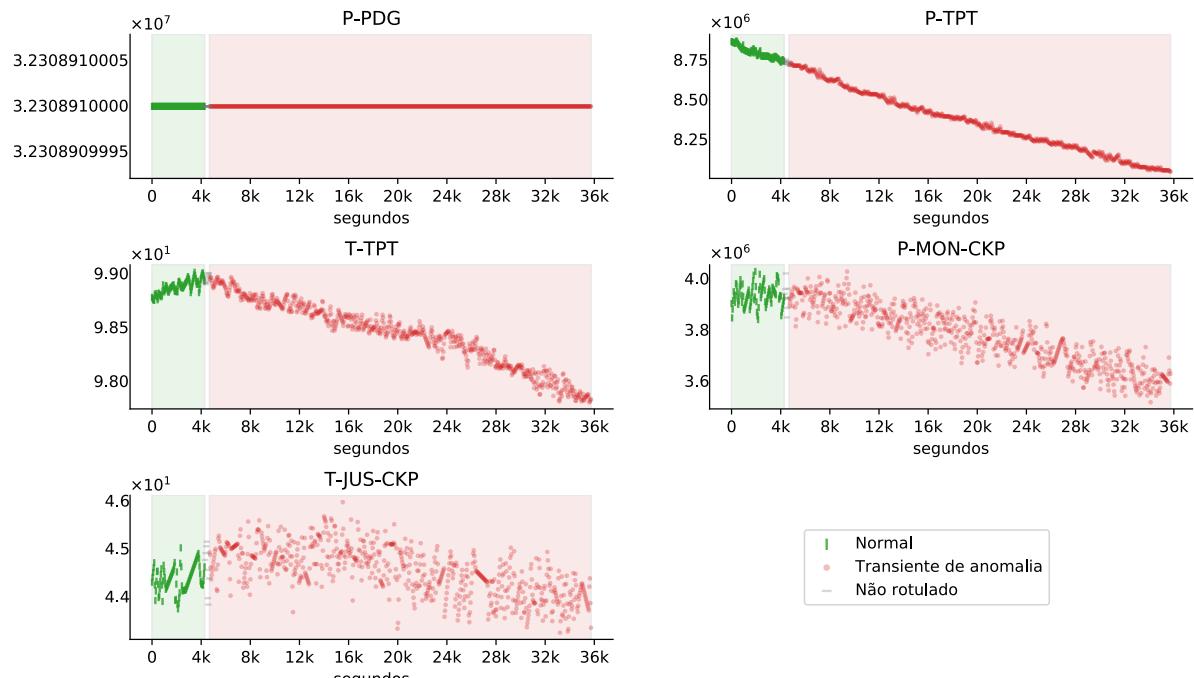
Como o reservatório em produção está em constante modificação ao ser esvaziado, é esperado que essas propriedades se alterem com o decorrer do tempo. Quando essas propriedades sofrem mudanças de tal forma que a energia do sistema deixa de ser suficiente para vencer as perdas, o fluido não consegue mais chegar até a superfície, cessando a produção. Esse fenômeno é conhecido na indústria de E&P como perda rápida de produtividade, que no limite significa perda de surgência.

Um prognóstico dessa condição permite a equipe de operação mudar o ponto de operação do poço para ele não perder a sua produtividade ou surgência, o que evita perda de produção.

Um exemplo de ocorrência real de perda de rápida de produtividade é exibido na Figura 7. Nesse caso, em cerca de 8 horas e 40 minutos após o início de transiência indicado por um especialista na área Elevação e Escoamento de Petróleo houve alguma atuação no poço e ele

não chegou ao estado estável desse tipo de anomalia. Novamente o P-PDG se encontrava congelado.

Figura 7 – Gráficos relacionados a um exemplo de ocorrência real de perda rápida de produtividade. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOF, 2019).

Nota: gráfico gerado pelo autor.

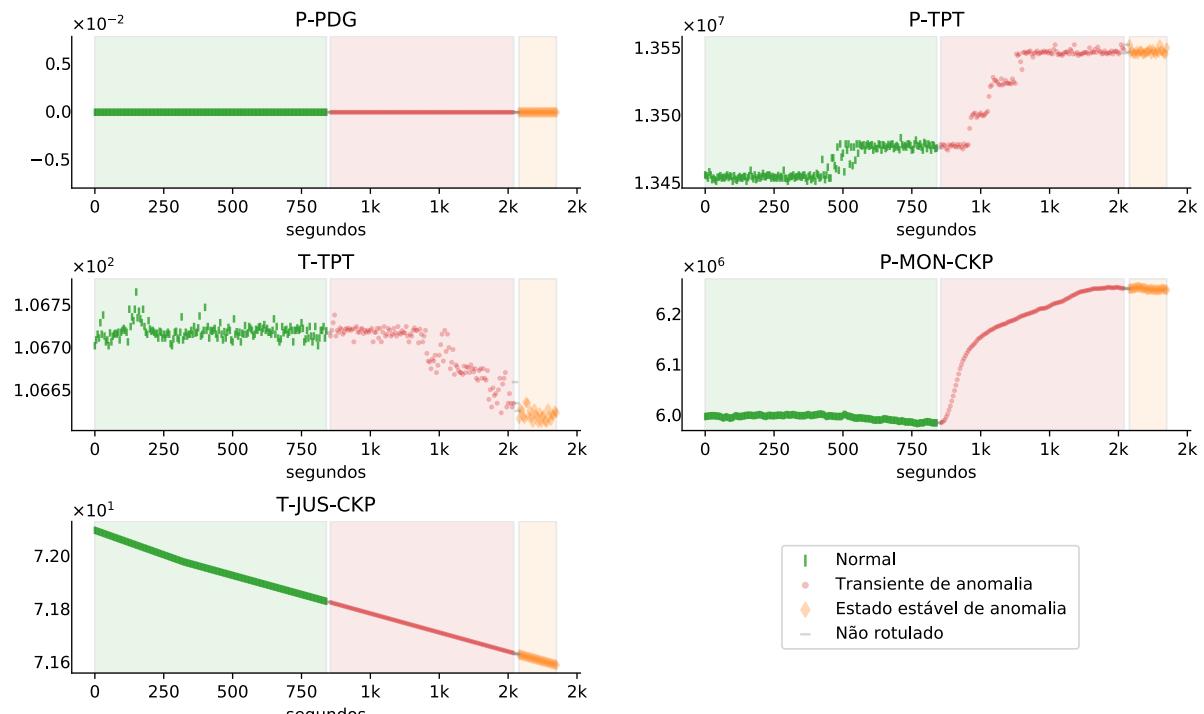
2.3.6 Restrição Rápida em CKP

A válvula CKP é instalada no início da unidade de produção, sendo responsável pelo controle do poço na superfície. A expressão “restrição rápida em CKP” não é bem definida na literatura, mas é corriqueiramente utilizada na Petrobras. Trata-se de uma restrição com amplitude acima de uma referência (5%, por exemplo) e durante um período restrito (até 10 segundos, por exemplo). Eventuais restrições rápidas indesejadas podem ocorrer nessa válvula por conta de problemas operacionais.

Ter um prognóstico desse tipo de anomalia também é desejável porque, em geral, essa válvula é manual e fechamentos indesejados podem ser desfeitos mais rapidamente, o que tende a diminuir perdas de produção.

A Figura 8 mostra as dinâmicas típicas das variáveis monitoradas quando da ocorrência de uma restrição rápida em válvula CKP. Nesse caso, o período de transiência de anomalia foi de, aproximadamente, 11 minutos.

Figura 8 – Gráficos relacionados a um exemplo de ocorrência real de restrição em CKP. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFT, 2019).

Nota: gráfico gerado pelo autor.

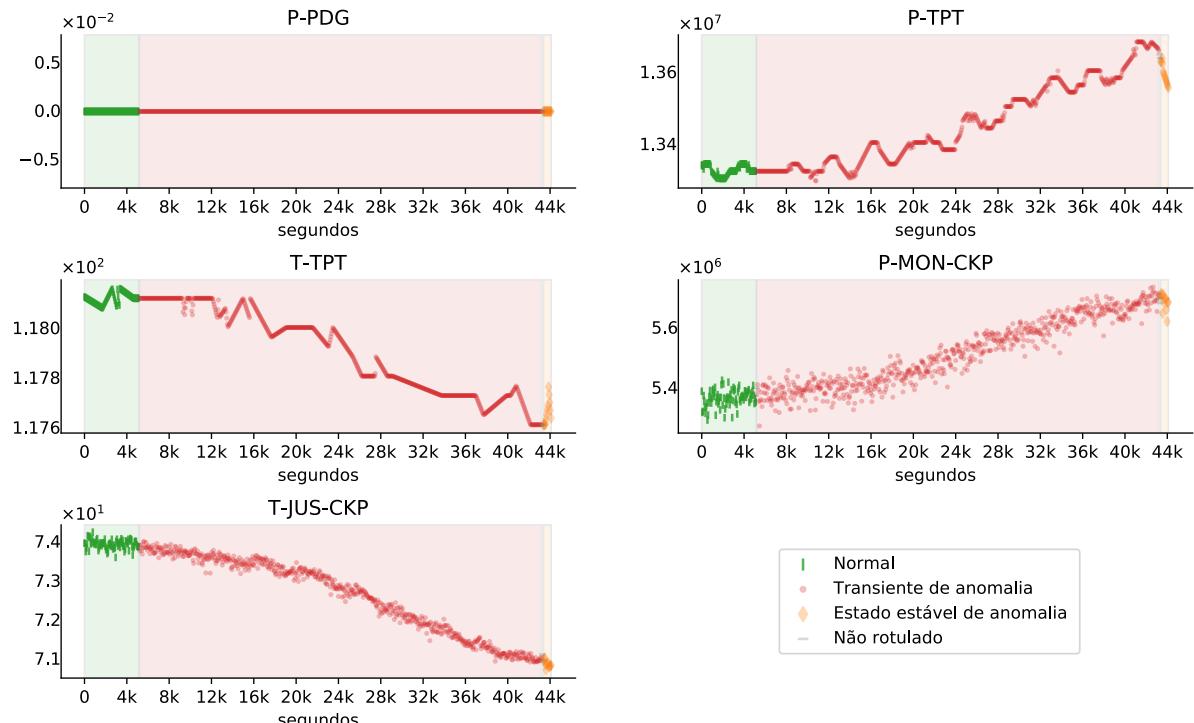
2.3.7 Incrustação em CKP

Monitoramento da válvula CKP é importante por conta da susceptibilidade de depósitos inorgânicos, que podem reduzir drasticamente a produção de petróleo (SCHLUMBERGER, 2019).

A identificação automática dessa condição em tempo hábil (prognóstico) também é desejável, pois ações apropriadas, como injeção de inibidor de incrustação, podem ser tomadas para evitar perdas de produção.

O exemplo de ocorrência real desse tipo de anomalia apresentado na Figura 9 teve um período de transiência de anomalia de, aproximadamente, 10 horas e 35 minutos.

Figura 9 – Gráficos relacionados a um exemplo de ocorrência real de incrustação em CKP. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFT, 2019).

Nota: gráfico gerado pelo autor.

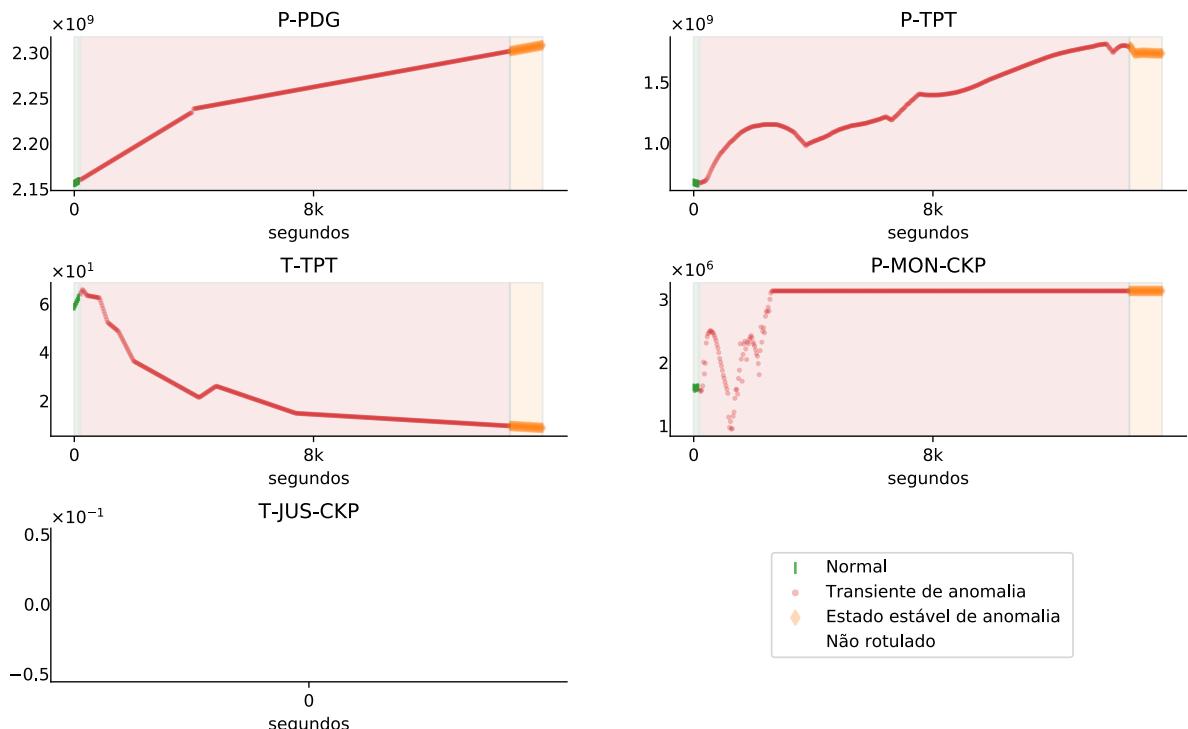
2.3.8 Hidrato em Linha de Produção

Hidrato é um dos grandes problemas na indústria de petróleo. Ele é definido como composto cristalino formado por água e gás natural e, por isso, se assemelha a gelo. Como a sua formação exige presença de água e gás natural, além de altas pressões e baixas temperaturas, oleodutos que escoam óleo morto não sofrem com este tipo de anomalia. Sua ocorrência se dá com mais frequência em gasodutos e poços produtores de gás. Contudo, hidratos também podem ser formados em poços produtores de óleo, chegando ao ponto de interromper totalmente suas vazões (ANDREOLLI, 2016) (ELLISON; GALLAGHER; LORIMER, 2000).

Evitar ocorrências deste tipo de anomalia significa evitar perdas de produção durante dias ou até semanas. Em certos casos, evita-se também custos altos para desobstrução da linha. Não é incomum, por exemplo, esse procedimento exigir sonda marítima cujo custo diário pode ultrapassar 500 mil dólares (ANDREOLLI, 2016).

Na Figura 10, pode-se ver um caso de hidrato em linha de produção cuja formação foi estimada em 4 horas. O decaimento da T-TPT indica que houve interrupção total da produção.

Figura 10 – Gráficos relacionados a um exemplo de ocorrência real de hidrato em linha de produção. P-PDG, P-TPT e P-MON-CKP em Pa. T-TPT e T-JUS-CKP em °C.



Fonte: PI System (OSISOFT, 2019).

Nota: gráfico gerado pelo autor.

2.4 TAMANHOS DE JANELAS TEMPORAIS DURANTE MONITORAMENTO

Para confirmar ocorrências reais de cada tipo de anomalia descrito na Seção 2.3, os profissionais que realizam monitoramento de poços na Petrobras normalmente analisam janelas temporais com tamanhos diferentes e capazes de compreenderem a dinâmica de cada respectivo tipo de anomalia. Estimativas para os tamanhos dessas janelas são apresentadas na Tabela 1.

Faz-se importante destacar que esses tamanhos de janelas, estimados por especialistas, podem ser utilizados para melhorar o desempenho de algoritmos de Aprendizado de Maquina, principalmente na fase de extração de características.

Tabela 1 – Estimativas de tamanhos de janelas temporais utilizados para confirmar ocorrências de anomalias.

TIPO DE ANOMALIA	TAMANHO DE JANELA
1 – Aumento Abrupto de BSW	12 h
2 – Fechamento Espúrio de DHSV	5 min – 20 min
3 – Intermitênciá Severa	5 h
4 – Instabilidade de Fluxo	15 min
5 – Perda Rápida de Produtividade	12 h
6 – Restrição Rápida em CKP	15 min
7 – Incrustação em CKP	72 h
8 – Hidrato em Linha de Produção	30 min – 5 h

Fonte: próprio autor.

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo reúne os principais fundamentos teóricos utilizados nesta tese.

3.1 FUNCIONAMENTO DOS ALGORITMOS DE APRENDIZADO DE MÁQUINA

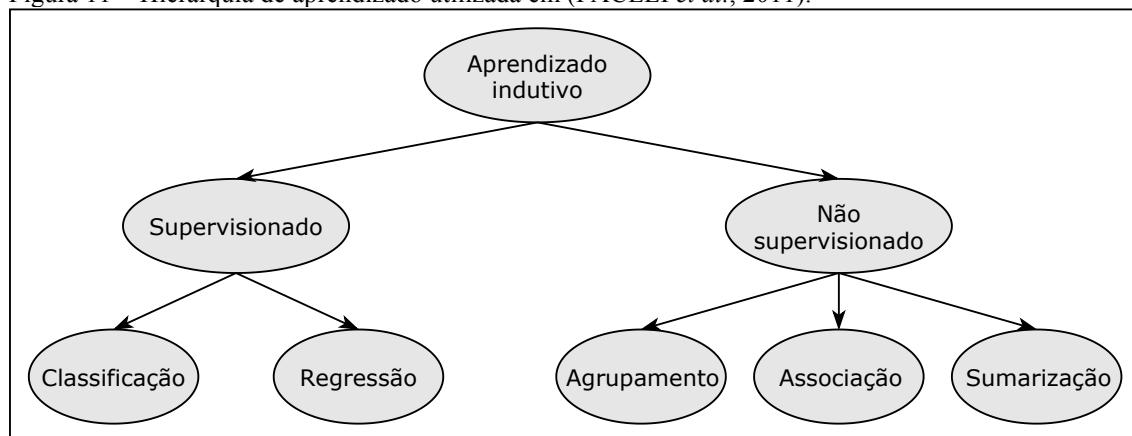
Em algoritmos de Aprendizado de Máquina, o aprendizado é obtido e armazenado em formato de hipótese, função ou modelo capaz de resolver um problema ou tarefa de interesse. Diz-se que o aprendizado (hipótese, função ou modelo) é induzido (ou generalizado) por esses algoritmos a partir de exemplos fornecidos, também chamados de objetos ou de instâncias. Então, esse aprendizado pode ser aplicado a outros exemplos para múltiplos propósitos (FACELI *et al.*, 2011).

Portanto, de uma forma geral, esses algoritmos funcionam em duas fases:

- Treinamento: fase inicial durante a qual o aprendizado acontece. Ou seja, é nessa fase que ocorre a indução de hipótese, função ou modelo;
- Utilização: fase em que o aprendizado é utilizado de fato para resolver o problema ou tarefa de interesse.

Esse conjunto de algoritmos pode ser organizado com vários critérios, o que resulta em diferentes hierarquias e quantidades de categorias e níveis de profundidade. Um critério muito utilizado, apresentado na Figura 11 e detalhado em (FACELI *et al.*, 2011), organiza os algoritmos de acordo com os seus paradigmas de aprendizado. Sobre este critério, é descrito a seguir apenas o estritamente necessário a esta contextualização.

Figura 11 – Hierarquia de aprendizado utilizada em (FACELI *et al.*, 2011).



Fonte: (FACELI *et al.*, 2011).

Nota: ilustração adaptada pelo autor.

Com o critério de Faceli *et al* (2011), as tarefas são inicialmente separadas entre preditivas (paradigma supervisionado) e descritivas (paradigma não-supervisionado).

Em uma tarefa de predição, os objetos utilizados precisam ter atributos de entrada e atributos de saída. Para cada objeto, as suas características são os seus atributos de entrada e os seus rótulos são seus atributos de saída. Um algoritmo dessa categoria induz um modelo a partir de características e rótulos associados a exemplos de treinamento para tentar prever rótulos associados a outros exemplos não utilizados no treinamento. O termo supervisionado é utilizado justamente para indicar a existência e a utilização de rótulos, cujos valores, em geral, são definidos por um “supervisor externo”. Cada característica ou rótulo pode ser discreto ou contínuo. Quando um algoritmo utiliza apenas rótulos discretos, ele é de classificação, o tipo mais relevante neste trabalho. Por outro lado, quando um algoritmo utiliza apenas rótulos contínuos, diz-se que ele é de regressão.

Em uma tarefa descritiva, rótulos não são utilizados, em geral por eles não estarem disponíveis. A meta desses algoritmos é explorar ou descrever um conjunto de objetos. Algoritmos de agrupamento visam encontrar grupos de objetos que têm características semelhantes. Por exemplo, na área de Marketing, tais algoritmos são utilizados para agrupamento de clientes em função de características específicas. Algoritmos de associação buscam encontrar regras entre grupos de características. Por exemplo, em estabelecimentos comerciais, esses algoritmos são utilizados para obtenção de regras do seguinte tipo: X% dos clientes que adquirem o produto A também adquirem na mesma compra o produto B, Y% dos clientes que adquirem os produtos C e D, devolvem o produto C em até Z dias. Por fim, há os algoritmos de summarização, que visam encontrar descrições simples e compactas para o conjunto de objetos fornecidos.

As características utilizadas por algoritmos de Aprendizado de Máquina podem ser fixas ou mutáveis no tempo.

A *Iris Data Set*, criada por Fisher (1936) e disponibilizada publicamente na Internet por Marshall (1988), talvez seja o exemplo mais clássico de base de dados e todas as suas características são fixas no tempo. Trata-se de um conjunto com 4 características, 1 rótulo com 3 possíveis valores (classes) e 50 instâncias reais associadas a cada classe (total de 150 instâncias). Cada classe se refere a um tipo diferente da planta Iris. Para cada objeto, nesse caso uma planta, suas 4 características não mudam com o passar do tempo. Como, em geral, os

objetos são independentes entre si, a ordem com a qual eles aparecem na base de dados não é relevante.

Um exemplo de base de dados com características mutáveis no tempo é a *Trace Data Set*, criada sinteticamente com *scripts* computacionais por Roverso (2000) e parcialmente disponibilizada publicamente na Internet por Chen *et al.* (2015). Essa base é composta por 4 características, 1 rótulo com 16 possíveis valores (classes) e 50 instâncias associadas a cada classe (total de 800 instâncias). Cada classe se refere a um tipo diferente de falha associada a instrumentação em uma usina nuclear. Cada característica é uma variável de processo cujos valores são obtidos de um sensor simulado. Nesse caso, cada objeto se refere a cada amostragem (leitura) dos sensores. Todas as características são, portanto, mutáveis no tempo (séries temporais).

O 3W dataset permite extração de características de ambos os tipos, fixas (origem da ocorrência e nome do poço) e mutáveis no tempo (estatísticas em geral, derivadas, coeficientes de Transformada *Wavelet*, coeficientes de Transformada de Fourier, número de observações, coeficientes de autocorrelação, quantidade de transições entre valores negativos/positivos ou vice-versa, etc.).

3.2 SÉRIE TEMPORAL MULTIVARIÁVEL

Neste trabalho, é adotada a notação de Série Temporal Multivariável (STM), do termo *Multivariate Time Series* (MTS), definida a seguir e similar às utilizadas em (ZHOU; CHAN, 2015) (WENG, 2013) (HE *et al.*, 2013).

Um *dataset* \mathbf{DS} é um conjunto de m MTS ($\mathbf{S}^i \mid i = 1, 2, \dots, m; m \in \mathbb{N} \text{ e } m > 1$), e é definido como $\mathbf{DS} = \{\mathbf{S}^1, \mathbf{S}^2, \dots, \mathbf{S}^m\}$. Cada MTS i é uma *instância* (também referenciada neste documento como objeto ou ocorrência), que é composta por um conjunto de n séries temporais univariáveis ($x_j^i \mid j = 1, 2, \dots, n; n \in \mathbb{N} \text{ e } n > 1$) (também referenciadas como *variáveis de processo* ou apenas *variáveis*), e é definida como $\mathbf{S}^i = \{x_1^i, x_2^i, \dots, x_n^i\}$. Cada variável j que compõe uma MTS i é uma sequência temporal ordenada de p_i observações extraídas no tempo t ($x_{j,t}^i \mid t = 1, 2, \dots, p_i; p_i \in \mathbb{N} \text{ e } p_i > 1$). Portanto, cada MTS i é vista neste trabalho como uma matriz definida como $\mathbf{S}^i = \{x_{1,1}^i, x_{2,1}^i, \dots, x_{n,1}^i; x_{1,2}^i, x_{2,2}^i, \dots, x_{n,2}^i; \dots; x_{1,p_i}^i, x_{2,p_i}^i, \dots, x_{n,p_i}^i\}$.

Note que todas as instâncias têm um número fixo de variáveis n , mas cada instância pode ser composta por qualquer quantidade de observações p_i . Também é importante notar que todas as variáveis de uma instância i têm número fixo de observações p_i .

O termo amostra utilizado neste documento se refere a uma MTS composta por observações obtidas entre dois instantes que delimitam uma janela temporal. Portanto, cada amostra também é uma MTS.

3.3 ALGORITMOS EXPERIMENTADOS

A seguir, são brevemente caracterizados os algoritmos experimentados nos *benchmarks* propostos neste trabalho para demonstração de que as hipóteses estabelecidas são verdadeiras. Leituras adicionais podem ser feitas nos trabalhos referenciados ou na documentação do *scikit-learn* (2019).

Um classificador do tipo *ensemble* combina resultados de múltiplos classificadores mais simples, chamados de fracos (*weak*), para realizar a sua tarefa. Aldrich e Auret (2013) definem classificadores fracos como aqueles que têm desempenho ligeiramente melhor que adivinhação aleatória. Classificações intermediárias são reunidas e, por meio de um critério pré-estabelecido, dão origem à classificação final. Dois dos possíveis critérios são: média aritmética simples e moda entre os resultados intermediários. Tecnicamente, não há limite para o número de classificadores fracos utilizados. Um classificador *ensemble* pode ser composto de centenas, milhares ou mais classificadores fracos. Diz-se que o modelo final desse tipo de algoritmo é o conjunto formado pelo modelo de cada classificador fraco utilizado.

Boosting algorithm é uma classe de algoritmos do tipo *ensemble* que obtém o seu modelo final de forma iterativa. Além disso, os resultados das classificações intermediárias são combinados de forma aditiva. A cada iteração, mais classificadores com modelos diferentes são adicionados no conjunto e mais ênfase é dada nas características mais difíceis de serem aprendidas. Ou seja, o desempenho do modelo final desse tipo de classificador é melhorado a cada iteração até um critério de parada especificado ser cumprido. Alguns dos possíveis critérios de parada são: número máximo de iterações e/ou limite inferior de melhoria no desempenho entre duas iterações consecutivas. Os classificadores fracos podem ser baseados em árvore, mas essa não é uma limitação desse conceito.

3.3.1 *Decision Tree*

Árvore de Decisão, ou *Decision Tree*, se refere a uma classe de algoritmos que particionam recursivamente o espaço de características para obterem subespaços com distribuição de saída cada vez mais pura. Sendo que a pureza é definida por alguma medida, por exemplo entropia de informação no caso de classificação e desvio médio quadrático no caso de regressão. (ALDRICH; AURET, 2013).

O nó raiz e os intermediários contêm regras de decisão aprendidas durante o treinamento. Já os nós folhas contêm, em uma tarefa de classificação, o rótulo da classe associada à respectiva partição do espaço de característica.

3.3.2 *AdaBoost*

Adaboost é um acrônimo para *Adaptive Boosting* (aprimoramento adaptativo), um *boosting algorithm* específico. Múltiplos classificadores fracos são aprimorados interativamente, por meio de ajustes em seus pesos, para classificarem corretamente os casos cada vez mais difíceis. O modelo final é obtido por votação dos classificadores fracos treinados, que por padrão são baseados em árvore de decisão (ZHU *et al.*, 2006).

Cada classificador fraco é treinado com um subconjunto aleatório das instâncias. A cada iteração, as instâncias que foram classificadas anteriormente com erro recebem pesos maiores. Assim, elas recebem maior ênfase no treinamento. Quanto maior é o desempenho de cada classificador fraco, maior é o peso do seu voto no modelo final.

3.3.3 *Random Forest*

Random Forest é um algoritmo do tipo *ensemble* no qual vários classificadores baseados em árvore de decisão, que constituem uma floresta, são treinados com subconjuntos distintos dos dados de treinamento. Cada subconjunto é obtido com seleção aleatória e com reposição de instâncias (ALDRICH; AURET, 2013).

Além da seleção com reposição, outra característica peculiar deste algoritmo é que durante o treinamento de cada árvore de decisão, a definição da regra de decisão a ser utilizada em cada nó é feita com apenas uma parte aleatória das características. Essa característica resulta em árvores menos correlacionadas entre si e isso tende a melhorar o desempenho do modelo, que é obtido por votação.

3.3.4 Isolation Forest

Esse algoritmo é único no sentido de isolar explicitamente anomalias em vez de criar modelos de instâncias normais. Assim como o *Random Forest*, este também é do tipo *ensemble* baseado em árvore de decisão (LIU; TING; ZHOU, 2008).

O primeiro passo desse algoritmo é obter uma amostra aleatória para treinamento. Particionamentos recursivos dessa amostra dão origens a uma árvore de decisão. Cada árvore é obtida com duas escolhas aleatórias em cada particionamento: característica e valor de corte entre os valores mínimo e máximo para a característica selecionada. A quantidade necessária de particionamentos para isolar uma instância é equivalente ao tamanho do caminho entre o nó raiz e a sua folha. O tamanho médio de caminho entre as árvores que compõem a floresta é uma medida de normalidade e, por isso, é utilizada como função de decisão.

Uma característica importante deste método é o fato dele ser treinado de forma não-supervisionada. Espera-se que o conjunto de treinamento seja composto por instâncias de uma única classe. Isso porque instâncias de outras classes (anomalias) no treinamento tendem a prejudicar a qualidade da função de decisão. Este método é, portanto, adequado para classificação de classe única (KRAWCZYK *et al.*, 2017).

3.3.5 Gradient Boosting

Gradient Boosting é um *boosting algorithm* específico que utiliza classificadores fracos baseados em árvore. O termo *gradient* compõe o seu nome porque a direção e o sentido do passo entre a iteração atual e a seguinte são obtidos por meio do gradiente da função de perda previamente escolhida. É justamente por meio desse gradiente que é dada, na iteração seguinte, mais ênfase nas características menos aprendidas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) (DUDA; HART; STORK, 2001).

Um resultado secundário do *Gradient Boosting* é a mensuração de relevância de cada característica no modelo final. Essa mensuração pode ser feita por meio de vários critérios. Uma forma clássica ordena as características pelo número de vezes em que cada uma delas foi utilizada pelos classificadores fracos. Essa mensuração de relevância pode ser utilizada como critério para seleção de características, por exemplo.

3.3.6 *Support Vector Machine*

Support Vector Machine (SVM) é uma família de algoritmos que buscam aprender hiperplanos, definidos por vetores de suporte, que otimizem a separação de duas classes. A sua versão mais simples obtém hiperplano linear no espaço de características que melhor distingue as classes. Outras versões mais sofisticadas se utilizam de *kernels*, que são funções não lineares que mapeia o espaço de características em outro de maior dimensionalidade no qual a separabilidade entre as classes tende a ser maior (ALDRICH; AURET, 2013) (WITTEN; FRANK; HALL, 2011).

Dentre os principais *kernels* existentes, são utilizados neste trabalho os seguintes: Polinomial (POLY), Sigmóide (SIGMOID) e a *Radial Base Function* (RBF).

3.3.7 *One Class SVM*

Este algoritmo da família SVM é próprio para classificação de classe única. Durante o treinamento, neste caso não-supervisionado, busca-se aprender uma função, linear ou não a depender do *kernel* utilizado, que retorne valores positivos para instâncias internas ao subconjunto de suporte ou, caso contrário, valores negativos (ALDRICH; AURET, 2013).

3.4 MÉTRICAS DE DESEMPENHO

São várias as métricas de desempenho existentes na literatura. Neste trabalho, são utilizadas as métricas aplicadas em (VARGAS; MUNARO; CIARELLI, 2016) com as seguintes adaptações relacionadas ao contexto:

- Verdadeiro-positivo (*VP*): número de amostras da classe *C* classificadas corretamente;
- Falso-negativo (*FN*): número de amostras da classe *C* classificadas equivocadamente;
- Verdadeiro-negativo (*VN*): número de amostras de outras classes diferentes de *C* classificadas corretamente (como diferente de *C*);
- Falso-positivo (*FP*): número de amostras de outras classes classificadas equivocamente (como *C*);
- Precisão (*P*): proporção entre o número de amostras da classe *C* estimadas como *C* e o número total de amostras estimadas como *C*. Ou seja, entre todas que foram estimadas como *C*, proporção das estimadas corretamente (Equação 3.1);

- Sensibilidade (S): proporção entre amostras da classe C estimadas como C e todas as amostras da classe C . Ou seja, entre todas as amostras da classe C , proporção das estimadas corretamente (Equação 3.2);
- Medida F ($F1$): média harmônica entre precisão e sensibilidade (Equação 3.3).

$$P = \frac{VP}{VP + FP} \quad (3.1)$$

$$S = \frac{VP}{VP + FN} \quad (3.2)$$

$$F1 = \frac{2 \times P \times S}{P + S} \quad (3.3)$$

Neste trabalho, espera-se que as amostras sejam classificadas com precisão e com sensibilidade. Por conta da Medida F incorporar essa relação de compromisso, ela foi escolhida como a principal métrica de desempenho neste trabalho.

4 METODOLOGIA

A metodologia aplicada na preparação do 3W dataset e na proposição dos *benchmarks* é detalhada neste capítulo. Ela foi concebida e desenvolvida no âmbito deste projeto de pesquisa e no contexto da Petrobras, levando-se em consideração as características gerais do problema descrito no Capítulo 2.

Não se defende que as escolhas feitas durante o desenvolvimento desta metodologia são ótimas. Contudo, argumenta-se que elas são razoáveis o suficiente e deram origem a contribuições relevantes: o 3W dataset em si e *benchmarks* que podem ser utilizados em conjunto com esse conjunto de dados.

4.1 PREPARAÇÃO DO 3W DATASET

O 3W dataset é composto por três tipos de instâncias que são determinados pela sua origem: real histórica, simulada ou desenhada à mão. Instâncias reais históricas, ou apenas reais, são aquelas que ocorreram de fato em poços produtores da Petrobras. Instâncias simuladas e desenhadas à mão são utilizadas fundamentalmente para diminuir o desbalanceamento do conjunto de dados inicialmente formado por apenas instâncias reais, o que é uma característica comum em dados industriais (HE *et al.*, 2013) (KRAWCZYK *et al.*, 2017).

Uma possível abordagem para diminuir esse tipo de desbalanceamento é realizar reamostragem entre as instâncias reais. Em (CHAWLA *et al.*, 2002), por exemplo, é proposto um método que realiza *over-sampling* da classe minoritária (eventos raros) e *under-sampling* da classe predominante (condição normal). A estratégia utilizada neste trabalho, entretanto, é completamente diferente. Busca-se enriquecer o conhecimento *a priori* (conjunto de dados) com mais instâncias obtidas em origens diferentes: simulações e curvas desenhadas à mão por especialistas no domínio do problema.

Rotulagens de dois tipos são realizadas por especialistas em cada anomalia. O primeiro é aplicado o nível da instância. Cada instância, independentemente se real histórica, simulada ou desenhada, foi necessariamente rotulada com um código numérico próprio associado à operação normal ou a algum outro código numérico atribuído à anomalia existente em algum instante desta instância. Note, portanto, que nenhuma instância tem mais de uma anomalia. Duas vantagens decorrentes desse tipo de rotulagem são que ele provê uma forma de agrupamento de instâncias em função do tipo de anomalia que elas contêm e que ele permite desenvolvimento

de classificadores *offline*, aqueles que não se propõem a estimar quando a anomalia iniciou ou terminou dentro da instância.

A rotulagem do segundo tipo foi aplicada no nível de observação. Cada observação de cada instância de qualquer tipo foi rotulada com um código próprio associado à operação normal ou a algum outro código numérico atribuído à anomalia existente naquele instante. Esse tipo de rotulagem é essencialmente destinado para treinamento de classificadores *online*.

Rotulagem no nível de observação foi realizada de forma a estabelecer até três períodos em cada instância de qualquer tipo: normal, transiente de anomalia e estado estável de anomalia. Um período no contexto deste trabalho significa uma sequência contígua de observações. Em períodos normais, não há evidência de nenhum tipo de anomalia. Em períodos transientes de anomalia, ou apenas transientes, a dinâmica resultante da anomalia ainda está em curso. Quando essa dinâmica cessa, um período estado estável de anomalia se inicia. O principal propósito dessa estratégia foi prover a possibilidade de prognóstico (classificação antecipada entre múltiplas classes). Ou seja, períodos transientes podem ser aprendidos e as suas corretas detecções predizem períodos estados estáveis de anomalias. Em outras palavras, o período transiente de anomalia pode ser interpretado como um período de pré-anomalia. A Figura 12 e a Figura 13 trazem exemplos de instâncias reais rotuladas como normal e anômala, respectivamente. No exemplo de normalidade, a variável T-TPT estava congelada. A anormalidade exemplificada é do tipo restrição rápida de CKP. Em ambos os casos, apenas três variáveis mais relevantes de acordo com especialistas são apresentadas. Diferentes tipos de períodos têm fundos com diferentes cores: verde para normal, amarelo para transiente de anomalia, vermelho para estado estável de anomalia e branco para observação não rotuladas (por conta de limitação da ferramenta computacional utilizada).

Todas as instâncias reais históricas foram extraídas do *plant information system* utilizado na UO-ES, o PI System (OSISOFT, 2019). Tal extração foi realizada com função disponível em *Application Programming Interface* (API) da própria OSISOFT que retorna série temporal periódica gerada com interpolação linear entre valores arquivados sem periodicidade definida. Para aspectos realistas serem mantidos no conjunto de dados, essa extração foi realizada com apenas um tipo de pré-processamento: substituição de valores inválidos (não numéricos) por valores ausentes. Assim, foram preservados os seguintes aspectos realistas: valores ausentes, variáveis congeladas (devido a problemas em sensores ou em rede de comunicação), instâncias com tamanhos diferentes entre si, *outliers* e diferentes níveis de exceção e compressão (que

representam filtros) aplicados pelo PI System nas diferentes variáveis (com possíveis reconfigurações ao longo do tempo). Essa estratégia permite avaliações sobre quais técnicas de pré-processamento resultam em desempenhos melhores em cada tipo de tarefa de interesse. Por outro lado, os processos utilizados para geração de instâncias simuladas e desenhadas à mão deram origem, naturalmente, a MTS livres desses tipos de problemas.

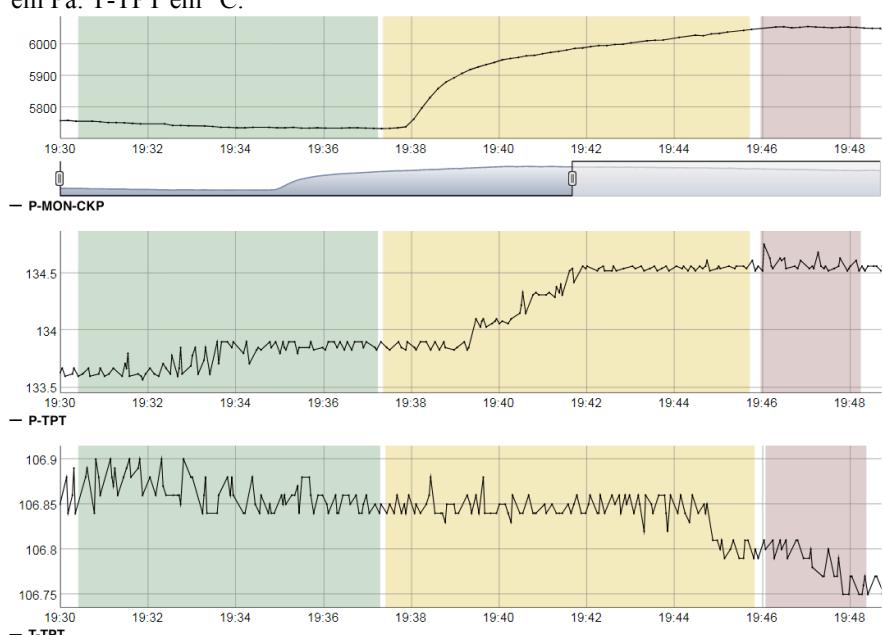
Figura 12 – Instância real histórica rotulada como normal. P-TPT e P-MON-CKP em Pa. T-TPT em °C.



Fonte: dados históricos da UO-ES.

Nota: gráfico gerado pelo autor.

Figura 13 – Instância real histórica rotulada como anômala. P-TPT e P-MON-CKP em Pa. T-TPT em °C.



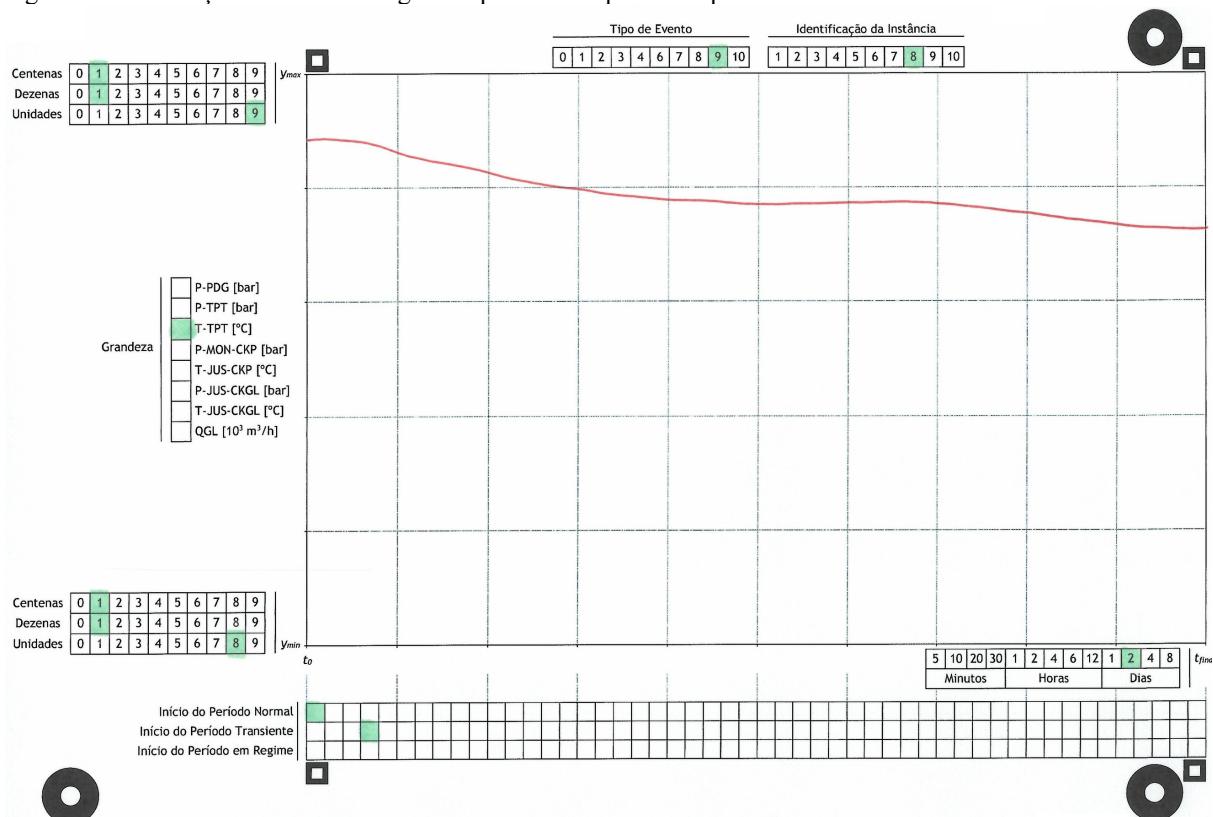
Fonte: dados históricos da UO-ES.

Nota: gráfico gerado pelo autor.

Todas as instâncias simuladas foram obtidas com o OLGA (SCHLUMBERGER, 2019), um simulador dinâmico multifásico adotado por diversas companhias ao redor do globo (ANDREOLLI, 2016) e modelos calibrados por especialistas na área Elevação e Escoamento de Petróleo. Essa escolha foi feita porque o OLGA é a ferramenta padrão na Petrobras para simulação de situações em poços de petróleo e pelo fato dele ser um dos poucos sistemas que simulam fenômenos dinâmicos (GRØDAHL, 2014). Simulações de anomalias reais raras foram priorizadas.

Uma ferramenta específica foi desenvolvida para enriquecer o 3W dataset com instâncias desenhadas à mão, que naturalmente incluem conhecimento tácito de especialistas em relação a formatos de MTS que caracterizam os tipos de anomalias considerados. Dois tipos raros de anomalias reais foram priorizados, um que foi simulado e outro que não. A ferramenta desenvolvida é composta por um modelo de gráfico e um *script* para processamento de imagem. A Figura 14 ilustra o uso desse modelo, no qual um especialista desenhou à mão e especificou todos os seus atributos: grandeza (variável); tipo de evento (anomalia); identificação da instância; início dos períodos normal, transiente de anomalia e estado estável de anomalia (em regime); e escalas.

Figura 14 – Ilustração do modelo de gráfico preenchido por um especialista na anomalia 9.



Fonte: especialista na anomalia 9.

Independentemente da origem, todas as instâncias foram geradas sempre com as mesmas variáveis. Além das cinco variáveis associadas aos sensores geralmente disponíveis em poços surgentes, outras três também foram consideradas. Isso por conta da falta de histórico que revelaria em quais momentos quais poços operaram via elevação natural ou algum outro método artificial. Essas três variáveis adicionais são ligadas a sensores posicionados na linha de serviço e podem ser úteis no prognóstico de anomalias que se manifestam elevação tanto natural quanto artificial. São elas:

- Pressão do fluido jusante à válvula *Choke de Gas Lift* (P-JUS-CKGL);
- Temperatura do fluido jusante à válvula *Choke de Gas Lift* (T-JUS-CKGL);
- Vazão de *Gas Lift* (QGL).

4.2 BENCHMARKS PROPOSTOS

Dois *benchmarks* específicos que praticantes e pesquisadores podem utilizar em conjunto com o 3W dataset foram concebidos como parte desta tese e proposto em (VARGAS *et al.*, 2019). O objetivo dessas propostas é prover uma forma padronizada e apropriada para algoritmos, implementados por participantes diversos com diferentes técnicas e abordagens, terem seus desempenhos avaliados e comparados.

Esses *benchmarks* foram planejados exclusivamente para detecção *online*. Todas as observações de períodos transientes de anomalia e estados estáveis de anomalia devem ser rotulados como positivos e todas as observações de períodos normais como negativas. Nessa operação, observações não rotuladas devem ser mantidas como estão.

Aspectos não incluídos nas regras desses *benchmarks* podem ser avaliados e escolhidos livremente. Alguns desses aspectos são técnicas de pré-processamentos, tamanho de janela temporal, abordagem para deslizamento de janela, número de amostras, abordagem para rotulagem de amostras e métodos para extração de características.

4.2.1 Impacto do Uso de Instâncias Simuladas e Desenhadas à Mão

Este *benchmark* foi desenvolvido para investigar o impacto do uso de instâncias simuladas e desenhadas à mão em treinamento de algoritmos de Aprendizado de Máquina para detecção de anomalias raras em instâncias reais. As regras a seguir devem ser observadas:

Regra 1: Apenas os tipos de anomalias 1 ou 7 devem ser escolhidos. Não é permitido mesclá-los na mesma classificação. Apenas esses tipos de anomalias têm instâncias das três origens (reais, simuladas e desenhadas à mão)⁶. Portanto, este *benchmark* impõe dois desafios específicos relacionados a classificação binária (detecção de anomalia).

Regra 2: Independentemente da origem, apenas instâncias rotuladas como do tipo escolhido podem ser utilizadas. Em outras palavras, apenas arquivos com extensão CSV salvos no diretório cujo nome é o tipo escolhido de anomalia podem ser utilizados.

Regra 3: Múltiplas rodadas de treinamento e validação devem ser realizadas com o esquema que chamamos de “deixe uma instância real de fora”. Faz-se importante destacar que esse esquema é diferente do clássico *leave-one-out* (BISHOP, 2006), que se fosse utilizado, a cada rodada apenas uma amostra seria deixada de lado. No esquema aqui imposto, todas as amostras provenientes da instância real deixada de fora não devem ser utilizadas no treinamento da rodada vigente. O número de rodadas deve ser o número de instâncias reais. Em cada rodada, sete cenários devem usar diferentes conjuntos de treinamento, mas precisamente o mesmo conjunto de validação composto apenas pela instância real deixada de fora do conjunto de treinamento deve ser utilizada. Todas as outras instâncias reais devem ser utilizadas no conjunto de treinamento. Cada instância real deve ser deixada de fora apenas uma vez. A partir da instância no conjunto de validação, o mesmo número de amostras de cada classe (positiva e negativa) devem ser extraídas. Cada cenário deve ser implementado de forma que o seu conjunto de treinamento seja composto por certo número de instância por origem, como descrito a seguir:

- Cenário 1: apenas instâncias reais;
- Cenário 2: apenas instâncias simuladas;
- Cenário 3: apenas instâncias desenhadas à mão;
- Cenário 4: apenas instâncias reais e simuladas;
- Cenário 5: apenas instâncias reais e desenhadas à mão;
- Cenário 6: apenas instâncias simuladas e desenhadas à mão;
- Cenário 7: instâncias reais, simuladas e desenhadas à mão.

⁶ Após a publicação deste *benchmark* em (VARGAS *et al.*, 2019) identificou-se que ele não é aplicável ao tipo de anomalia 7, pois na verdade a versão vigente do 3W dataset não tem instâncias simuladas desse tipo (vide Tabela 2).

Regra 4: Em cada rodada, precisão, sensibilidade e medida F devem ser computadas, mas outras métricas também podem ser consideradas. Valor médio e desvio padrão de cada métrica entre todas as rodadas devem ser apresentados. Valor médio da medida F deve ser considerado a principal métrica de desempenho, por estabelecer uma relação de compromisso entre precisão e sensibilidade.

4.2.2 Detecção de Anomalia

A intenção deste *benchmark* é encorajar desenvolvimento, avaliações e comparações de algoritmos que realizam detecção de anomalia. Nessa tarefa, anomalias devem ser distinguidas de condições normais. As regras a seguir devem ser observadas:

Regra 1: Apenas instâncias reais com anomalias de tipos que têm períodos normais (1, 2, 5, 6, 7 e 8) maiores ou iguais a vinte minutos devem ser utilizadas. Aquelas com rótulos diferentes não podem ser utilizadas. Em outras palavras, apenas arquivos com extensão CSV salvos em diretório cujo nome é um desses tipos podem ser utilizados.

Regra 2: Múltiplas rodadas de treinamento e validação devem ser realizadas. O número de rodadas deve ser igual ao número de instâncias. Em cada rodada, o seguinte cenário deve ser implementado. Amostras utilizadas para treinamento ou validação devem ser extraídas de apenas uma instância. Parte das amostras negativas devem ser utilizadas no treinamento e a outra parte na validação. Todas as amostras positivas devem ser utilizadas apenas na validação. Portanto, uma técnica de aprendizagem de classe única deve ser utilizada. O conjunto de validação deve ser composto pelo mesmo número de amostras de cada classe (positiva e negativa).

Regra 3: Em cada rodada, precisão, sensibilidade e medida F devem ser computadas, mas outras métricas também podem ser consideradas. Valor médio e desvio padrão de cada métrica entre todas as rodadas devem ser apresentados. Valor médio da medida F deve ser considerado a principal métrica de desempenho, por estabelecer uma relação de compromisso entre precisão e sensibilidade.

5 RESULTADOS

A seguir são apresentados os resultados associados às contribuições científicas relevantes desta tese. Tratam-se do 3W dataset em si e dos resultados obtidos com experimentações relacionadas aos *benchmarks* propostos.

Tais experimentações se encontram integralmente disponíveis em (VARGAS *et al.*, 2019). Elas foram estruturadas com o Jupyter Notebook (PROJECT JUPYTER, 2019) e os seus códigos foram desenvolvidos com a linguagem Python (PYTHON SOFTWARE FOUNDATION, 2019).

Faz-se importante citar que outros *scripts* precisaram ser desenvolvidos neste trabalho, inclusive para obtenção, manipulação e tratamento de séries temporais provenientes tanto de ocorrências reais quanto de instâncias simuladas e desenhadas à mão. Todos eles também foram desenvolvidos com a linguagem Python.

5.1 3W DATASET

O 3W dataset foi estruturado da forma descrita a seguir e se encontra disponível em (VARGAS *et al.*, 2019, 2019a).

Cada instância, independentemente se real, simulada ou desenhada à mão, foi salva em um arquivo dedicado a ela, padronizado e com extensão *Comma-Separated Values* (CSV). Todos esses arquivos foram agrupados em diretórios baseados em rótulo no nível de instância. Todas as instâncias foram geradas com observações obtidas com taxa de amostragem fixa (1 Hz). Apenas as unidades de engenharia a seguir foram utilizadas: Pascal [Pa], metros cúbicos por segundo padronizados [sm^3/s] e graus Celsius [$^\circ\text{C}$]. A origem de cada instância foi incorporada no nome do seu arquivo. Os nomes de todos os poços foram substituídos por nomes genéricos como requisito da Petrobras para publicação do 3W dataset.

A Tabela 2 apresenta as quantidades de instâncias que compõem o 3W dataset por tipo de instância e por origem de conhecimento: reais, simuladas e desenhadas à mão. Considerando-se o limiar de 1% e apenas instâncias reais, quatro tipos de anomalias são raros: códigos 1, 6, 7 e 8. Se instâncias simuladas também são consideradas, três desses tipos deixam de ser raros: códigos 1, 6 e 8. Mesmo após considerar as instâncias desenhadas à mão, um tipo de anomalia ainda permanece raro: código 7.

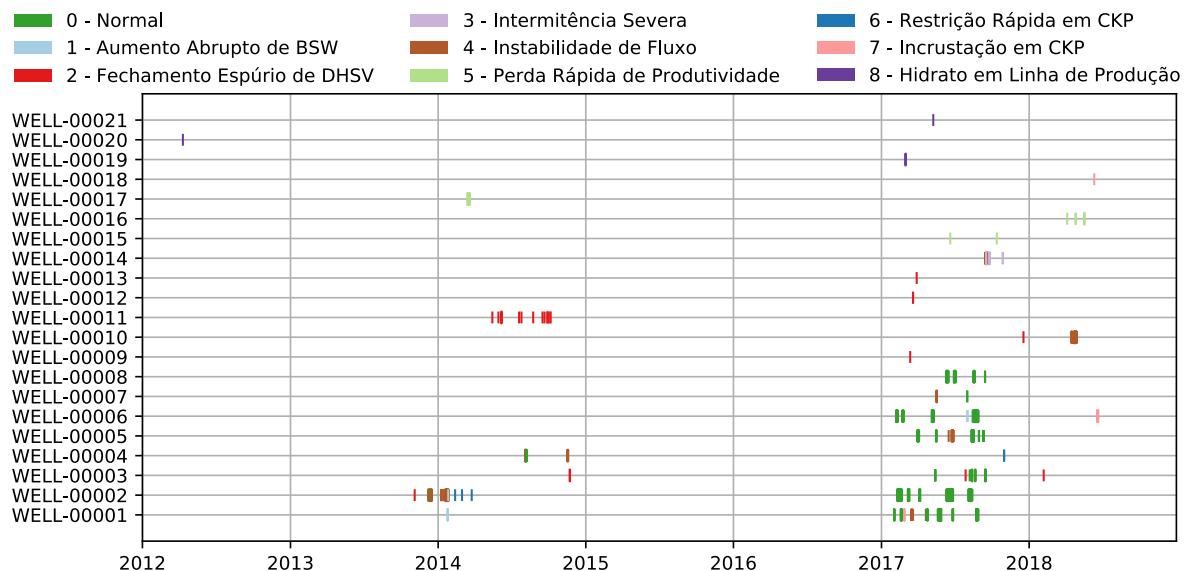
Tabela 2 – Quantidades de instâncias que compõem o 3W dataset.

TIPO DE INSTÂNCIA	REAIAS	SIMULADAS	DESENHADAS	TOTAL
	INSTÂNCIAS			
0 – Normal	597	–	–	597
1 – Aumento Abrupto de BSW	5	114	10	129
2 – Fechamento Espúrio de DHSV	22	16	–	38
3 – Intermitência Severa	32	74	–	106
4 – Instabilidade de Fluxo	344	–	–	344
5 – Perda Rápida de Produtividade	12	439	–	451
6 – Restrição Rápida em CKP	6	215	–	221
7 – Incrustação em CKP	4	–	10	14
8 – Hidrato em Linha de Produção	3	81	–	84
TOTAL	1025	939	20	1984

Fonte: próprio autor.

A Figura 15 apresenta um mapa de dispersão das instâncias reais históricas contidas no 3W dataset. A mais antiga ocorreu em meados de 2012 e a mais recente em meados de 2018. Além da quantidade de poços reais considerados (21), esse mapa também provê uma visão geral das distribuições das ocorrências de cada tipo de instância (normal e anomalias) tanto ao longo do tempo, quanto entre poços.

Figura 15 – Mapa de dispersão das instâncias reais históricas do 3W dataset.



Fonte: próprio autor.

Os principais aspectos do 3W dataset relacionados a dificuldades inerentes a dados reais são:

- Variáveis ausentes: 4.947 (31,17% de todas as 15.872 variáveis de todas as 1.984 instâncias). Quando todas as observações de uma variável em uma instância particular

têm valores ausentes por conta de problemas em sensor ou rede de comunicação, essa variável em si é considerada ausente. Quanto mais variáveis ausentes, o *dataset* se torna mais esparso, o que pode impor dificuldades adicionais aos algoritmos;

- Variáveis congeladas: 1.535 (9,67% de todas as 15.872 variáveis de todas as 1.984 instâncias). Quando todas as observações de uma variável em particular têm qualquer valor real ou inteiro único por conta de qualquer motivo, essa variável em si é considerada congelada. Uma variável congelada nem sempre representa um problema, mas essa característica é usualmente sintoma de problema em sensor, em configuração de sistema ou em rede de comunicação. Portanto, variáveis congeladas problemáticas não manifestam os padrões associados às anomalias, o que de fato impõe dificuldades adicionais aos algoritmos;
- Observações não rotuladas: 5.130 (0,01% de todas as 50.913.215 observações de todas as 15.872 variáveis de todas as 1.984 instâncias). Algumas observações não foram rotuladas por conta de limitações da ferramenta utilizada. Mesmo com esse percentual, técnicas podem ser avaliadas para tratamento das observações não rotuladas.

5.2 IMPACTO DO USO DE INSTÂNCIAS SIMULADAS E DESENHADAS À MÃO

Além das quatro regras impostas por este *benchmark* e descritas na Subseção 4.2.1, as seguintes escolhas foram feitas nesta experimentação:

- Reamostragem de uma a cada 60 observações em todas as instâncias em função da dinâmica relativamente lenta do tipo de anomalia 1 (vide Tabela 1). Esta escolha visa reduzir o tempo de execução desta experimentação sem comprometer significativamente os seus resultados;
- Hiperparâmetros não foram otimizados, pois o objetivo desta experimentação é gerar uma base para comparações com resultados de algoritmos especializados que venham a ser desenvolvidos. Para cada hiperparâmetro, foi utilizado o valor padrão definido no *scikit-learn* (2019);
- Uma estratégia específica de amostragem com janela deslizante foi utilizada para cada tipo de período em instâncias de treinamento. Acredita-se que essa estratégia facilita a aprendizagem das características das anomalias. A saber:
 - Em períodos normais, as últimas observações foram privilegiadas para aprendizagem de características mais próximas de períodos transientes de

- anomalia. Até 15 amostras com 5 observações cada (após reamostragem) foram utilizadas;
- Em períodos transientes, buscou-se utilizar as observações ao longo do período como um todo. Até 15 amostras com 60 observações cada (após reamostragem) foram utilizadas;
 - Em períodos estados estáveis de anomalias, as primeiras observações foram privilegiadas para aprendizagem de características mais próximas de períodos transientes. Até 15 amostras com 60 observações cada (após reamostragem) foram utilizadas;
 - Independentemente do tipo de período, as amostras foram extraídas possivelmente com sobreposição. Esse recurso foi utilizado nos períodos com tamanhos insuficientes para extração de todas as amostras sem sobreposição. Nesses casos, buscou-se utilizar a menor sobreposição possível.
- A estratégia de amostragem em instâncias de validação foi diferente. Afinal informações sobre os períodos não podem ser utilizadas. Buscou-se utilizar as observações ao longo de cada instância como um todo. Até 45 amostras com 60 observações cada (após reamostragem) foram extraídas de cada instância de validação;
 - Antes de cada rodada de treinamento e validação:
 - As instâncias utilizadas foram normalizadas de forma que todas as variáveis passaram a ter média zero e variância unitária. Caso contrário, a diferença entre as escalas poderia prejudicar os resultados da experimentação;
 - As variáveis de instâncias reais utilizadas no treinamento que tinham quantidades de valores ausentes acima de um limiar (10%) ou que tinham desvios padrões abaixo de outro limiar (1%) foram totalmente descartadas no treinamento e na validação. Cada valor ausente remanescente foi substituído por zero. Entende-se que essa estratégia é simples e efetiva para seleção de variáveis.
 - Para reduzir o tempo de execução desta experimentação, foi utilizado um conjunto mínimo de características: mediana, média, desvio padrão, variância, máximo e mínimo.

Foram realizadas cinco rodadas de treinamento e validação, mesmo número de instâncias reais do tipo de anomalia aqui considerado conforme estabelecido na regra 3⁷. Em cada rodada, os

⁷ A regra 3 foi parcialmente respeitada. Em cada rodada, o número de amostras de cada classe (positiva e negativa) extraídas da instância de validação nem sempre foi o mesmo. Esse relaxamento se mostrou necessário devido parte das instâncias reais do tipo de anomalia selecionado terem períodos normais menores do que a sua dinâmica

quatro algoritmos citados na Tabela 4 foram treinados e validados em cada um dos sete cenários definidos no *benchmark*. Portanto, ao todo foram realizadas 140 ($5 \times 4 \times 7$) avaliações e em cada uma delas foram calculadas as métricas solicitadas.

A Tabela 3 agrupa as médias e os desvios padrões (entre parênteses) das métricas em cada cenário independentemente do algoritmo. Cada média e desvio padrão foram estimados, portanto, entre 20 valores (métricas). A dispersão dos valores associados a cada uma dessas métricas é apresentada da Figura 16 à Figura 19.

Já a Tabela 4 apresenta as médias e os desvios padrões (entre parênteses) das métricas para cada algoritmo independentemente do cenário. Nesse caso, cada média e desvio padrão foram calculados entre 35 valores (métricas). A dispersão dos valores associados a cada uma dessas métricas é apresentada da Figura 20 a Figura 21.

Tanto a Tabela 3 quanto a Tabela 4 estão ordenadas decrescentemente em função das médias de F1. Percebe-se que o cenário 1 resultou nas piores métricas, tanto em relação a média quanto desvio padrão. Já o desempenho do algoritmo *Random Forest* foi numericamente superior aos dos demais, analisando-se média e desvio padrão de qualquer uma das três métricas.

Tabela 3 – Médias e desvio padrão das métricas consideradas, por cenário, na experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas. Foram obtidos valores iguais para as três métricas devido ao uso do parâmetro “*average='micro'*” (SCIKIT-LEARN, 2019).

CENÁRIO	F1	PRECISÃO	SENSIBILIDADE
5	0,904 (0,131)	0,904 (0,131)	0,904 (0,131)
4	0,883 (0,222)	0,883 (0,222)	0,883 (0,222)
2	0,849 (0,138)	0,849 (0,138)	0,849 (0,138)
6	0,799 (0,230)	0,799 (0,230)	0,799 (0,230)
7	0,760 (0,300)	0,760 (0,300)	0,760 (0,300)
3	0,607 (0,368)	0,607 (0,368)	0,607 (0,368)
1	0,499 (0,406)	0,499 (0,406)	0,499 (0,406)

Fonte: próprio autor.

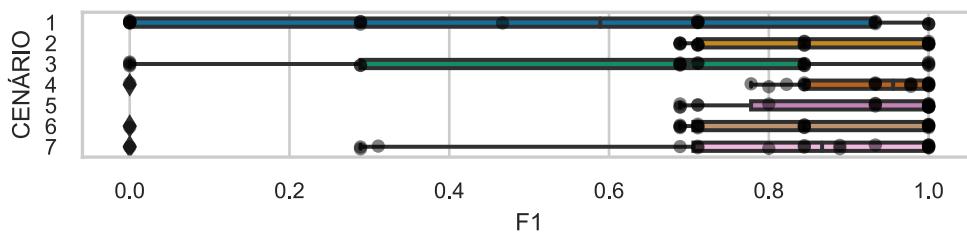
temporal típica. Se amostras menores fossem extraídas, elas não aprenderiam essa dinâmica. Optou-se, portanto, em extrair amostras com tamanho adequado em relação a essa dinâmica a partir de cada instância de validação. Para respeitar o regramento relacionado ao número de rodadas, o relaxamento recém explicado se fez obrigatório.

Tabela 4 – Médias e desvio padrão das métricas consideradas, por algoritmo, na experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas. Foram obtidos valores iguais para as três métricas devido ao uso do parâmetro “average=’micro’” (SCIKIT-LEARN, 2019).

ALGORITMO	F1	PRECISÃO	SENSIBILIDADE
<i>Random Forest</i>	0,825 (0,249)	0,825 (0,249)	0,825 (0,249)
<i>Gradient Boosting</i>	0,787 (0,267)	0,787 (0,267)	0,787 (0,267)
<i>AdaBoost</i>	0,785 (0,269)	0,785 (0,269)	0,785 (0,269)
<i>Decision Tree</i>	0,632 (0,384)	0,632 (0,384)	0,632 (0,384)

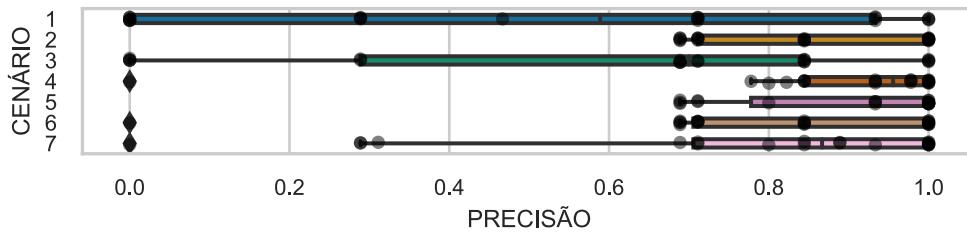
Fonte: próprio autor.

Figura 16 – Dispersão da F1, por cenário, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.



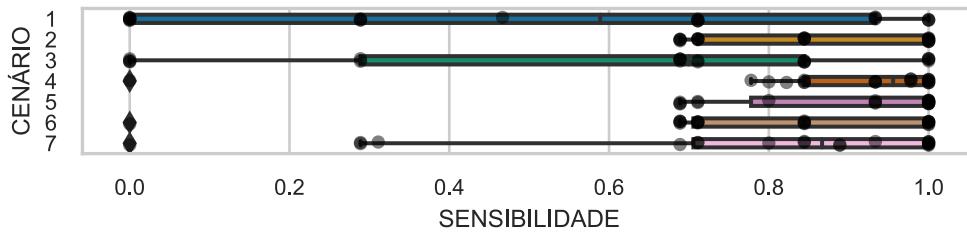
Fonte: próprio autor.

Figura 17 – Dispersão da precisão, por cenário, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.



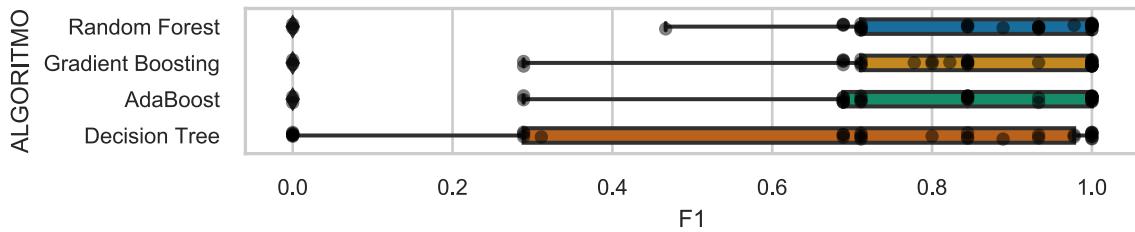
Fonte: próprio autor.

Figura 18 – Dispersão da sensibilidade, por cenário, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.



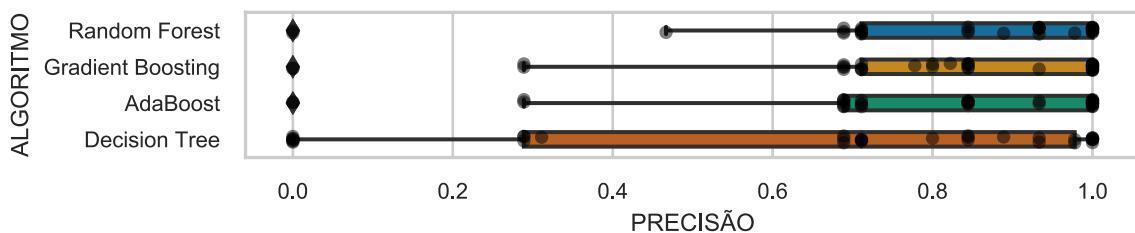
Fonte: próprio autor.

Figura 19 – Dispersão da F1, por algoritmo, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.



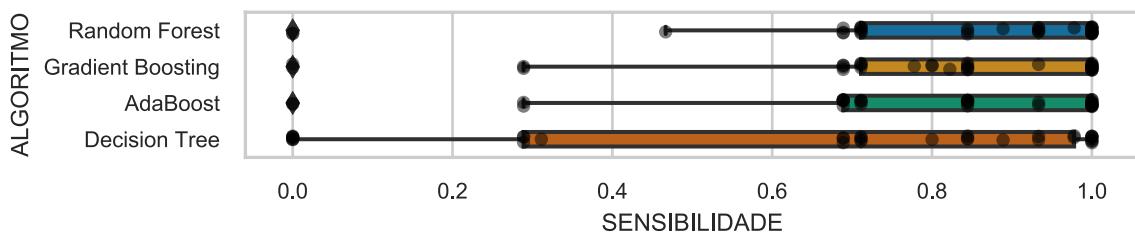
Fonte: próprio autor.

Figura 20 – Dispersão da precisão, por algoritmo, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.



Fonte: próprio autor.

Figura 21 – Dispersão da sensibilidade, por algoritmo, nos resultados da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.



Fonte: próprio autor.

Testes estatísticos foram aplicados para verificação se as médias obtidas (e quais), tanto para cada cenário quanto para cada algoritmo, são diferentes entre si com alta probabilidade. Inicialmente, no nível de bloco foi utilizado o Teste de Friedman. Nos casos em que a sua hipótese nula pode ser rejeitada, utilizou-se na sequência o Teste de Holm com grupo de controle e Correção de Bonferroni. Todos os testes estatísticos realizados são avaliados com significância de 5%. Especificamente os Testes de Friedman e de Holm foram escolhidos por serem não-paramétricos e em função da argumentação feita em (DEMSAR, 2006). Foram utilizadas implementações desses testes desenvolvidas por (RODRÍGUEZ-FDEZ *et al.*, 2015) e disponíveis em (RODRÍGUEZ-FDEZ *et al.*, 2015a).

Como o Teste de Friedman aplicado aos vetores de métrica F1 associados aos cenários resultou em um *p-value* igual a $1,7 \times 10^{-3}$, a hipótese de todos os cenários terem médias de F1 iguais

pode ser rejeitada. O resultado do Teste de Holm com o cenário 1 como de controle e Correção de Bonferroni é apresentado na Tabela 5. Esse resultado indica que apenas os cenários 4 e 5 têm, com alta probabilidade, médias diferentes estatisticamente em relação à média do cenário 1, no qual apenas instâncias reais foram utilizadas no conjunto de treinamento. Como no conjunto de treinamento dos cenários 4 e 5 foram utilizadas apenas, respectivamente, instâncias reais em conjunto com simuladas e instâncias reais em conjunto com desenhadas à mão, pode-se dizer que o uso de instâncias simuladas e desenhadas à mão elevou o desempenho médio dos detectores de instâncias reais da anomalia do tipo 1, que é raro.

Tabela 5 – Resultados do Teste de Holm no contexto dos cenários da experimentação relacionada a impacto do uso de instâncias simuladas e desenhadas.

COMPARAÇÃO	P-VALUE	P-VALUE CORRIGIDO
1 versus 4	$6,7 \times 10^{-4}$	$4,0 \times 10^{-3}$
1 versus 5	$1,3 \times 10^{-3}$	$6,4 \times 10^{-3}$
1 versus 2	$6,7 \times 10^{-2}$	$2,7 \times 10^{-1}$
1 versus 7	$9,2 \times 10^{-2}$	$2,8 \times 10^{-1}$
1 versus 6	$1,2 \times 10^{-1}$	$2,8 \times 10^{-1}$
1 versus 3	$7,4 \times 10^{-1}$	$7,4 \times 10^{-1}$

Fonte: próprio autor.

Já o Teste de Friedman aplicado aos vetores de métrica F1 associados aos algoritmos experimentados (e não aos cenários) resultou em um *p-value* igual a $2,8 \times 10^{-1}$, o que indica que a hipótese dos seus desempenhos médios (vide Tabela 4) serem iguais não pode ser rejeitada.

A Tabela 6 traz os principais resultados das experimentações realizadas para este *benchmark*.

Tabela 6 – Principais resultados das experimentações para o *benchmark* sobre o impacto do uso de instâncias simuladas e desenhadas à mão.

RESULTADO	SIGNIFICADO
Na média e em relação ao cenário 1, os algoritmos experimentados geraram resultados estatisticamente melhores apenas nos cenários 4 e 5.	A segunda hipótese desta tese é verdadeira.
Não se pode afirmar que os resultados médios dos algoritmos experimentados em todos os cenários são diferentes entre si.	Os algoritmos experimentados têm desempenhos similares para este <i>benchmark</i> .

Fonte: próprio autor.

5.3 DETECÇÃO DE ANOMALIA

Além das três regras impostas por este *benchmark* e descritas na Subseção 4.2.2, as seguintes escolhas foram feitas nesta experimentação:

- Hiperparâmetros não foram otimizados, pois o objetivo desta experimentação também é gerar uma base para comparações com resultados de algoritmos especializados que venham a ser desenvolvidos;
- Uma estratégia específica de amostragem com janela deslizante foi utilizada para cada tipo de período. Acredita-se que essa estratégia facilita a aprendizagem das características das anomalias. A saber:
 - Em períodos normais, as primeiras observações foram utilizadas para treinamento (60%) e as últimas para validação (40%). Até 15 amostras com 180 observações cada foram utilizadas;
 - Em períodos transientes, buscou-se utilizar as observações ao longo do período como um todo para validação (apenas). Até 15 amostras com 180 observações cada foram utilizadas;
 - Em períodos estados estáveis de anomalias, as primeiras observações foram privilegiadas para validação (apenas). Até 15 amostras com 180 observações cada foram utilizadas;
 - Independentemente do tipo de período, as amostras foram extraídas possivelmente com sobreposição. Esse recurso foi utilizado nos períodos com tamanhos insuficientes para extração de todas as amostras sem sobreposição. Nesses casos, buscou-se utilizar a menor sobreposição possível.
- Antes de cada rodada de treinamento e validação:
 - As observações utilizadas (e não instâncias) foram normalizadas de forma que todas as variáveis passaram a ter média zero e variância unitária. Caso contrário, a diferença entre as escalas poderia prejudicar os resultados da experimentação;
 - As variáveis das amostras (e não das instâncias) utilizadas no treinamento que tinham quantidades de valores ausentes acima de um limiar (10%) ou que tinham desvios padrões abaixo de outro limiar (1%) foram totalmente descartadas no treinamento e na validação. Cada valor ausente remanescente foi substituído por zero. Entende-se que essa estratégia é simples e efetiva para seleção de variáveis.
- Para reduzir o tempo de execução desta experimentação, foi utilizado um conjunto mínimo de características: mediana, média, desvio padrão, variância, máximo e mínimo.

Foram realizadas 36 rodadas de treinamento e validação, mesmo número de instâncias reais do tipo de anomalia aqui considerado conforme estabelecido na regra 2. Em cada rodada, os seis

algoritmos citados na Tabela 7 foram treinados e validados. Portanto, ao todo foram realizadas 216 (36×6) avaliações e em cada uma delas foram calculadas as métricas solicitadas.

A Tabela 7 agrupa as médias e os desvios padrões (entre parênteses) das métricas para cada algoritmo e está ordenada decrescentemente em função das médias de F1. Cada média e desvio padrão foram calculados, portanto, entre 36 valores (métricas). A dispersão dos valores associados a cada uma dessas métricas é apresentada da Figura 22 à Figura 24.

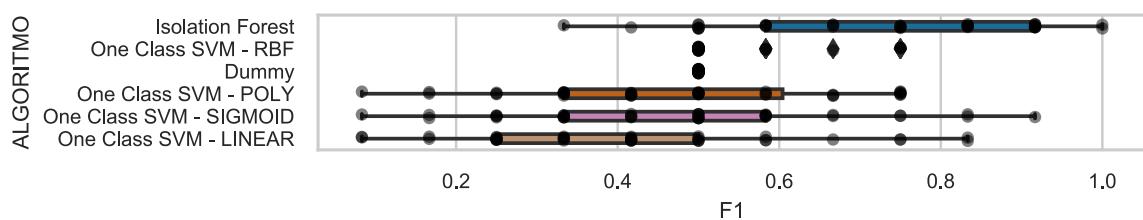
Percebe-se que o desempenho do algoritmo *Isolation Forest* é numericamente superior aos dos demais em relação às médias das métricas apresentadas, mas essa conclusão é diferente em relação aos seus desvios padrões.

Tabela 7 – Médias e desvio padrão das métricas consideradas, por algoritmo, na experimentação relacionada a detecção de anomalia. Foram obtidos valores iguais para as três métricas devido ao uso do parâmetro “average=’micro’” (SCIKIT-LEARN, 2019).

ALGORITMO	F1	PRECISÃO	SENSIBILIDADE
<i>Isolation Forest</i>	0,727 (0,182)	0,727 (0,182)	0,727 (0,182)
<i>One Class SVM - RBF</i>	0,532 (0,075)	0,532 (0,075)	0,532 (0,075)
<i>Dummy</i>	0,500 (0,000)	0,500 (0,000)	0,500 (0,000)
<i>One Class SVM - POLY</i>	0,472 (0,188)	0,472 (0,188)	0,472 (0,188)
<i>One Class SVM - SIGMOID</i>	0,470 (0,201)	0,470 (0,201)	0,470 (0,201)
<i>One Class SVM - LINEAR</i>	0,414 (0,195)	0,414 (0,195)	0,414 (0,195)

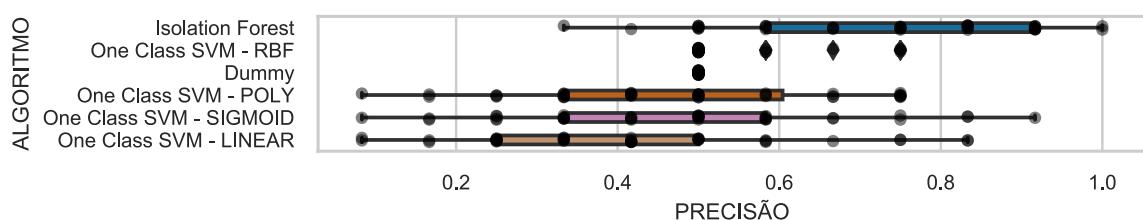
Fonte: próprio autor.

Figura 22 – Dispersão da F1 nos resultados da experimentação relacionada a detecção de anomalia.



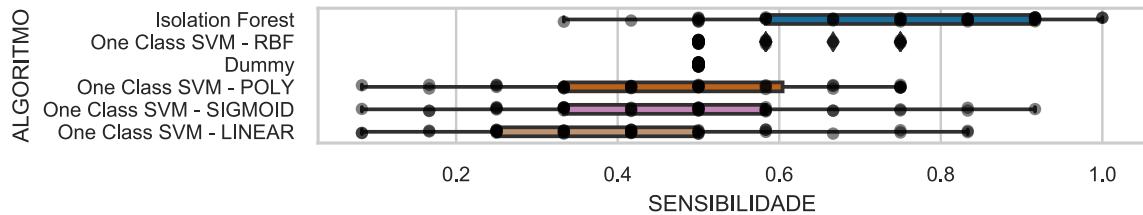
Fonte: próprio autor.

Figura 23 – Dispersão da precisão nos resultados da experimentação relacionada a detecção de anomalia.



Fonte: próprio autor.

Figura 24 – Dispersão da sensibilidade nos resultados da experimentação relacionada a detecção de anomalia.



Fonte: próprio autor.

Os mesmos testes estatísticos e significância estatística considerados na primeira experimentação são reutilizados nesta.

Como o Teste de Friedman aplicado aos vetores de métrica F1 associados aos algoritmos resultou em um $p\text{-value}$ igual a $1,4 \times 10^{-9}$, a hipótese de todos os algoritmos terem médias de F1 iguais pode ser rejeitada. O resultado do Teste de Holm com o algoritmo *Dummy*⁸ (ingênuo) como de controle e Correção de Bonferroni é apresentado na Tabela 8. Esse resultado indica que apenas o algoritmo *Isolation Forest* tem, com alta probabilidade, média diferente estatisticamente em relação à média do algoritmo ingênuo. Com isso, pode-se dizer que anomalias em poços de petróleo podem ser detectadas com Aprendizado de Máquina, em comparação a uma abordagem ingênua.

Tabela 8 – Resultados do Teste de Holm no contexto dos algoritmos da experimentação relacionada a detecção de anomalia.

COMPARAÇÃO	P-VALUE	P-VALUE CORRIGIDO
Dummy versus <i>Isolation Forest</i>	$3,2 \times 10^{-5}$	$1,6 \times 10^{-4}$
Dummy versus <i>One Class SVM - LINEAR</i>	$4,1 \times 10^{-2}$	$1,6 \times 10^{-1}$
Dummy versus <i>One Class SVM - SIGMOID</i>	$5,3 \times 10^{-1}$	$1,0 \times 10^0$
Dummy versus <i>One Class SVM - RBF</i>	$5,5 \times 10^{-1}$	$1,0 \times 10^0$
Dummy versus <i>One Class SVM - SIGMOID</i>	$5,7 \times 10^{-1}$	$1,0 \times 10^0$

Fonte: próprio autor.

A Tabela 9 traz os principais resultados das experimentações realizadas para este *benchmark*.

Tabela 9 – Principais resultados das experimentações para o *benchmark* sobre detecção de anomalia.

RESULTADO	SIGNIFICADO
Na média, o algoritmo <i>Isolation Forest</i> gerou resultado estatisticamente melhor em relação ao algoritmo <i>Dummy</i> .	A primeira hipótese desta tese é verdadeira.

Fonte: próprio autor.

⁸ Considerou-se como algoritmo ingênuo aquele que sempre prediz a classe predominante, ou seja, normalidade.

6 CONCLUSÕES

O problema de engenharia que motivou esta tese, ou seja, prognóstico de anomalias em sistemas de elevação natural de petróleo, foi caracterizado. Poços que operam com esse tipo de sistema, seus sensores geralmente disponíveis e as suas anomalias mais críticas em termos de potencial de impactar na produção de petróleo foram descritos.

Como etapa necessária para evidenciar a veracidade das hipóteses formuladas nesta tese, uma metodologia para preparação de conjunto de dados foi concebida e aplicada no contexto da Petrobras. Dessa aplicação surgiu o 3W dataset, até onde se sabe o primeiro *dataset* público com dados de anomalias reais em poços de petróleo, que pode ser utilizado como um conjunto de dados de referência para desenvolvimento de (i) técnicas de Aprendizado de Máquina para tratamento de dificuldades inerentes a dados reais e (ii) métodos para tarefas específicas associadas a classificação de anomalias em sistemas de elevação natural de petróleo.

Desafios específicos (*benchmarks*) foram propostos, nos quais algoritmos disponíveis publicamente foram experimentados e os resultados obtidos, mesmo com restrições em relação a otimização de hiperparâmetros, evidenciam que ambas as hipóteses formuladas são verdadeiras. Ou seja, anomalias em poços de petróleo podem ser detectadas com Aprendizado de Máquina e treinar algoritmos de Aprendizado de Máquina com exemplos de anomalias gerados por simulador computacional ou desenhados à mão por especialistas é uma solução viável para detecção de ocorrências reais raras dessas anomalias. Esses resultados obtidos também servem de base para comparações com resultados de algoritmos especializados que venham a ser desenvolvidos para tais desafios.

O 3W dataset em si e os *benchmarks* propostos representam as contribuições científicas relevantes deste trabalho.

Como trabalho futuro, sugerem-se:

- Explorar os *benchmarks* propostos, desenvolver algoritmos especializados e comparar os seus resultados com os resultados de base aqui apresentados;
- Desenvolver técnicas de Aprendizado de Máquina com o 3W dataset para tratamento de dificuldades inerentes a dados reais, por exemplo em termos de pré-processamento, filtros, transformações, família de classificadores, otimização de hiperparâmetros, engenharia de características e métricas de desempenho;

- Desenvolver métodos com o 3W dataset para tarefas não experimentadas nesta tese, por exemplo detecção de novas falhas, detecção e classificação *offline*, classificação de múltiplas classes e prognóstico (classificação antecipada).

REFERÊNCIAS

- ABASS, H.; BASS, D. **The Critical Production Rate in Water-Coning System**. Permian Basin Oil and Gas Recovery Conference. Texas: Society of Petroleum Engineers. 1988. p. 351-360.
- AHMADI, R.; AMINSHAHIDY, B.; SHAHRABI, J. Well-testing model identification using time-series shapelets. **Journal of Petroleum Science and Engineering**, v. 149, p. 292–305, 2017.
- ALDRICH, C.; AURET, L. **Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods**. London: Springer, 2013.
- ANDREOLLI, I. **Introdução à Elevação e Escoamento Monofásico e Multifásico de Petróleo**. Rio de Janeiro: Interciência, 2016.
- ARRUDA, F. et al. **Fault Detection in Industrial Plant Using k-Nearest Neighbors with Random Subspace Method**. International Conference on Artificial Intelligence. Las Vegas: Anais do International Conference on Artificial Intelligence. 2014.
- BAGNALL, A. et al. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. **Data Mining and Knowledge Discovery**, v. 31, n. 3, p. 606–660, 2017.
- BAGNALL, A. et al. **The UEA & UCR Time Series Classification Repository**, 2019. Disponível em: <www.timeseriesclassification.com>. Último acesso: 25 de julho de 2019.
- BHATTACHARYA, S.; MISHRA, S. Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: Case studies from the Appalachian basin, USA. **Journal of Petroleum Science and Engineering**, v. 170, p. 1005–1017, 2018.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. 1. ed. New York: Springer-Verlag, 2006. 738 p.
- CAMARGO, E. et al. **Intelligent Supervision Systems for Improving the Industrial Production Performance in Oil Wells**. International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics. Merida: Anais do International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics. 2010. p. 289-296.
- CAMPOS, M.; GOMES, M.; PEREZ, J. **Controle Avançado e Otimização na Indústria do Petróleo**. Rio de Janeiro: Interciência Ltda., 2013.
- C-FER TECHNOLOGIES. **C-FER Technologies**, 2019. Disponível em: <<https://www.cfertech.com>>. Último acesso: 25 de julho de 2019.
- C-FER TECHNOLOGIES. **Electric Submersible Pump - Reliability Information and Failure Tracking System (ESP-RIFTS)**, 2019. Disponível em: <<http://jip.esprifts.com>>. Último acesso: 29 de julho de 2019.

- CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002.
- CHEN, X. **Awesome Public Datasets**, 2019. Disponível em: <<https://github.com/awesomedata/awesome-public-datasets>>. Último acesso: 25 de julho de 2019.
- CHEN, Y. et al. **The UCR Time Series Classification Archive**, 2015. Disponível em: <www.cs.ucr.edu/~eamonn/time_series_data/>. Último acesso: 29 de julho de 2019.
- DAU, H. A. et al. **The UCR Time Series Classification Archive**, 2018. Disponível em: <www.cs.ucr.edu/~eamonn/time_series_data_2018/>. Último acesso: 25 de julho de 2019.
- DEMSAR, J. Statistical Comparisons of Classifiers over Multiple Data Sets. **Journal of Machine Learning Research**, v. 7, p. 1-30, 2006.
- DHEERU, D.; CASEY, G. **[UCI] Machine Learning Repository**, 2019. Disponível em: <<http://archive.ics.uci.edu/ml>>. Último acesso: 25 de julho de 2019.
- DONNEZ, P. **Essentials of Reservoir Engineering**. Paris: Editions Technip, 2007.
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2nd. ed. New York: Wiley-Interscience, 2001.
- ELLISON, B.; GALLAGHER, C.; LORIMER, S. **The Physical Chemistry of Wax, Hydrates, and Asphaltene**. Offshore Technology Conference. Houston: Anais do Offshore Technology Conference. 2000.
- FACELI, K. et al. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. Rio de Janeiro: LTC, 2011.
- FISHER, R. **The Use of Multiple Measurements in Taxonomic Problems**. Annual Eugenics. London: Blackwell Publishing Ltd. 1936. p. 179-188.
- GAING, Z.-L. Wavelet-based neural network for power disturbance recognition and classification. **IEEE Transactions on Power Delivery**, v. 19, n. 4, p. 1560–1568, 2004.
- GEURTS, P. **Pattern Extraction for Time Series Classification**. 5th European Conference on Principles of Data Mining and Knowledge Discovery. Freiburg: Springer, Berlin, Heidelberg. 2001. p. 115-127.
- GRØDAHL, S. I. **Small Scale Multiphase Flow Experiments on Surge Waves in Horizontal Pipes**. Norwegian University. Trondheim. 2014.
- GRAY, M.; MORSI, W. G. **Application of wavelet-based classification in non-intrusive load monitoring**. Canadian Conference on Electrical and Computer Engineering. Halifax: IEEE. 2015. p. 41–45.
- HARIDY, S.; WU, Z. Univariate and multivariate control charts for monitoring dynamic-behavior processes: A case study. **Journal of Industrial Engineering and Management**, p. 464-498, 2009.

- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. 2. ed. New York: Springer-Verlag, 2009.
- HAUSLER, R. H.; KRISHNAMURTHY, R. M.; SHERAR, B. **Observation of Productivity Loss in Large Oil Wells due to Scale Formation without Apparent Production of Formation Brine**. CORROSION 2015. Dallas: NACE International. 2015.
- HE, G. et al. **Early prediction on imbalanced multivariate time series**. International Conference on Information and Knowledge Management. San Francisco: Anais do International Conference on Information and Knowledge Management. 2013. p. 1889–1892.
- HE, Q. P.; WANG, J. Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes. **IEEE Transactions on Semiconductor Manufacturing**, Tucson, v. 20, n. 4, p. 345-354, 2007.
- ISTITUTO NAZIONALE DI STATISTICA. **Time Series**, 2019. Disponível em: <http://seriestoriche.istat.it/index.php?id=1&no_cache=1&L=1&no_cache=1>. Último acesso: 29 de julho de 2019.
- JAMES, G. et al. **An Introduction to Statistical Learning with Applications in R**. New York: Springer-Verlag, 2013.
- JOUNELA, S.-L. J. Future trends in process automation. **Annual Reviews in Control**, v. 31, p. 211–220, 2007.
- KRAWCZYK, B. et al. Ensemble learning for data stream analysis: A survey. **Information Fusion**, v. 37, p. 132–156, 2017.
- LI, D. et al. **Time Series Classification with Discrete Wavelet Transformed Data: Insights from an Empirical Study**. 28th International Conference on Software Engineering and Knowledge Engineering. Redwood City: Anais do 28th International Conference on Software Engineering and Knowledge Engineering. 2016.
- LI, K.; HAN, Y.; WANG, T. A novel prediction method for down-hole working conditions of the beam pumping unit based on 8-directions chain codes and online sequential extreme learning machine. **Journal of Petroleum Science and Engineering**, v. 160, p. 285–301, 2018.
- LIU, C.; LI, Y.; XU, M. An integrated detection and location model for leakages in liquid pipelines. **Journal of Petroleum Science and Engineering**, v. 175, p. 852–867, 2019.
- LIU, F. T.; TING, K. M.; ZHOU, Z. **Isolation Forest**. 2008 Eighth IEEE International Conference on Data Mining. Pisa: IEEE. 2008. p. 413-422.
- LIU, Y. et al. **Failure Prediction for Artificial Lift Systems**. SPE Western Regional Meeting. Anaheim: Society of Petroleum Engineers. 2010.
- LIU, Y. et al. **Failure Prediction for Rod Pump Artificial Lift Systems**. SPE Western Regional Meeting. Anaheim: Society of Petroleum Engineers. 2010.
- LIU, Y. et al. **Semi-supervised Failure Prediction for Oil Production Wells**. 11th International Conference on Data Mining Workshops. Vancouver: IEEE. 2011. p. 434-441.

- MARSHALL, M. Iris Data Set. **Machine Learning Repository**, 1988. Disponível em: <<https://archive.ics.uci.edu/ml/datasets/iris>>. Último acesso: 29 de julho de 2019.
- MEGLIO, F. et al. Stabilization of slugging in oil production facilities with or without upstream pressure sensors. **Journal of Process Control**, v. 22, n. 4, p. 809–822, 2012.
- OLMOS, R.; TABIK, S.; HERRERA, F. Automatic Handgun Detection Alarm in Videos Using Deep Learning. **Neurocomputing**, v. 275, n. C, p. 66–72, 2018.
- OSISOFT. PI System. **OSIsoft Web Site**, 2019. Disponível em: <<https://www.osisoft.com/pi-system/>>. Último acesso: 29 de julho de 2019.
- PATRI, O. et al. **Extracting Discriminative Shapelets from Heterogeneous Sensor Data**. International Conference on Big Data (Big Data). Washington: IEEE. 2014. p. 1095-1104.
- PATRI, O. et al. **Predicting Failures from Oilfield Sensor Data using Time Series Shapelets**. SPE Annual Technical Conference and Exhibition. Amsterdam: Society of Petroleum Engineers. 2014.
- PATRI, O. et al. **Multivariate Time Series Classification Using Inter-leaved Shapelets**. NIPS Time Series Workshop. Montreal: Neural Information Processing Systems. 2015.
- PATRI, O. et al. **Predicting Compressor Valve Failures from Multi-Sensor Data**. SPE Western Regional Meeting. Garden Grove: Society of Petroleum Engineers. 2015.
- PATRI, O. et al. **Data Mining with Shapelets for Predicting Valve Failures in Gas Compressors**. SPE Western Regional Meeting. Anchorage: Society of Petroleum Engineers. 2016.
- PIANTANIDA, M. et al. **Advanced Flow Assurance Diagnostics in Val D'agri**. Offshore Mediterranean Conference. Ravenna: Anais do Offshore Mediterranean Conference. 2013.
- PROJECT JUPYTER. **Project Jupyter Web Site**, 2019. Disponível em: <<http://jupyter.org>>. Último acesso: 31 de julho de 2019.
- PYTHON SOFTWARE FOUNDATION. **Python Web Site**, 2019. Disponível em: <<https://www.python.org>>. Último acesso: 31 de julho de 2019.
- ROCHA, L.; AZEVEDO, C. **Projetos de Poços de Petróleo**. 2a. ed. Rio de Janeiro: Editora Interciência, 2009. ISBN 9788571932142.
- RODRÍGUEZ-FDEZ, I. et al. **STAC Python Library**, 2015. Disponível em: <<http://tec.citius.usc.es/stac/doc/index.html>>. Último acesso: 09 de agosto de 2019.
- RODRÍGUEZ-FDEZ, I. et al. **STAC: A web platform for the comparison of algorithms using statistical tests**. International Conference on Fuzzy Systems (FUZZ-IEEE). Istanbul: IEEE. 2015. p. 1-8.
- ROSA, A. J.; CARVALHO, R. D. S.; XAVIER, J. A. D. **Engenharia de Reservatórios de Petróleo**. 1a. ed. Rio de Janeiro: Editora Interciência, 2006. ISBN 8571931356.

- ROVERSO, D. **Multivariate Temporal Classification By Windowed Wavelet Decomposition And Recurrent Neural Networks**. International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies. Washington: Anais do International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies. 2000.
- RUSSELL, E. L.; CHIANG, L. H.; BRAATZ, R. D. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 51, n. 1, p. 81–93, 2000.
- SANTOS, I. H. et al. **Hydrate Failure Detection In Production And Injection Lines Using Model And Data-Driven Approaches**. Rio Oil&Gas Expo and Conference. Rio de Janeiro: Anais do Rio Oil&Gas Expo and Conference. 2018.
- SCHLUMBERGER. OLGA Dynamic Multiphase Flow Simulator. **Schlumberger Web Site**, 2019. Disponível em: <<https://www.software.slb.com/products/olga>>. Último acesso: 29 de julho de 2019.
- SCHLUMBERGER. The Oilfield Glossary: Where the Oil Field Meets the Dictionary. **Schlumberger Web Site**, 2019. Disponível em: <<https://www.glossary.oilfield.slb.com>>. Último acesso: 29 de julho de 2019.
- SCHMIDT, Z.; DOTY, D.; DUTTA-ROY, K. Severe Slugging in Offshore Pipeline Riser-Pipe Systems. **Society of Petroleum Engineers Journal**, v. 25, n. 1, p. 27-38, fev. 1985.
- SCIKIT-LEARN. **scikit-learn - Machine Learning in Python**, 2019. Disponível em: <<http://scikit-learn.org>>. Último acesso: 27 de julho de 2019.
- STANDARDS NORWAY. **NORSOK Standard D-010**. Standards Norway. Lysaker. 2013.
- SUTHERLAND, V. A. et al. **Executive Summary – Drilling Rig Explosion and Fire at the Macondo Well**. Washington. 2016.
- TAKEI, J. et al. **Flow Instability In Deepwater Flowlines And Risers – A Case Study Of Subsea Oil Production From Chinguetti Field, Mauritania**. SPE Asia Pacific Oil and Gas Conference and Exhibition. Brisbane: Society of Petroleum Engineers. 2010.
- TANG, H. et al. Time series data analysis for automatic flow influx detection during drilling. **Journal of Petroleum Science and Engineering**, v. 172, p. 1103–1111, 2019.
- THEYAB, M. Severe Slugging Control: Simulation of Real Case Study. **Journal of Environmental Research**, v. 2, n. 1, 2018.
- THOMAS, J. E. **Fundamentos de Engenharia de Petróleo**. 2a. ed. Rio de Janeiro: Editora Interciência, 2004. ISBN 8571930996.
- UNG, H. et al. Intracranial EEG fluctuates over months after implanting electrodes in human brain. **Journal of Neural Engineering**, v. 14, n. 5, 2017.
- VARGAS, R. et al. **Proposal for Two Classifiers of Offshore Naturally Flowing Wells Events Using K-Nearest Neighbors, Sliding Windows and Time Multiscale**. 6th

- International Symposium on Advanced Control of Industrial Processes (AdCONIP). Taipei: IEEE. 2017. p. 209-214.
- VARGAS, R. et al. A realistic and public dataset with rare undesirable real events in oil wells. **Journal of Petroleum Science and Engineering**, v. 181, p. 106223, 2019.
- VARGAS, R. et al. Data for: A Realistic and Public Dataset with Rare Undesirable Real Events in Oil Wells. **Mendeley Data**, 2019. Disponível em: <<http://dx.doi.org/10.17632/r7774rwc7v.1>>. Último acesso: 25 de julho de 2019. Versão 1.
- VARGAS, R. et al. The first realistic and public dataset with rare undesirable real events in oil wells. **GitHub**, 2019. Disponível em: <https://github.com/ricardovvargas/3w_dataset>. Último acesso: 29 de julho de 2019.
- VARGAS, R.; MUNARO, C.; CIARELLI, P. **Um Método para Detecção de Causalidade de Granger com Seleção de Regressores**. XXI Congresso Brasileiro de Automática. Vitória: Sociedade Brasileira de Automática. 2016. p. 3397-3402.
- VENKATASUBRAMANIAN, V. et al. A Review of Process Fault Detection and Diagnosis: Part I. Quantitative Model-Based Methods. **Computer and Chemical Engineering**, v. 27, n. 3, p. 293–311, 2003.
- WENG, X. **Classification of Multivariate Time Series Using Supervised Neighborhood Preserving Embedding**. Chinese Control and Decision Conference (CCDC). Guiyang: IEEE. 2013. p. 957–961.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**. Third. ed. Burlington,: Morgan Kaufmann, 2011. 664 p.
- XAVIER, G. M.; SEIXAS, J. M. D. **Fault Detection and Diagnosis in a Chemical Process using Long Short-Term Memory Recurrent Neural Network**. International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro: Anais do International Joint Conference on Neural Networks (IJCNN). 2018. p. 1–8.
- XI, X. et al. **Fast Time Series Classification on Using Numerosity Reduction**. 23rd International Conference on Machine learning. Pittsburgh: Anais do 23rd International Conference on Machine learning. 2006. p. 1033-1040.
- XING, Z. et al. **Extracting Interpretable Features for Early Classification on Time Series**. 11th SIAM International Conference on Data Mining. Mesa: Anais do 11th SIAM International Conference on Data Mining. 2011. p. 247-258.
- XING, Z.; PEI, J.; YU, P. **Early Prediction on Time Series: A Nearest Neighbor Approach**. 21st International Joint Conference on Artificial Intelligence. Pasadena: Anais do 21st International Joint Conference on Artificial Intelligence. 2009. p. 1297-1302.
- ZHOU, P.-Y.; CHAN, K. **A Feature Extraction Method for Multivariate Time Series Classification Using Temporal Patterns**. 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Ho Chi Minh City: Springer, Cham. 2015. p. 409–421.
- ZHU, J. et al. Multi-class AdaBoost. **Statistics and its interface**, v. 2, 2006.