RODRIGO SCORALICK FONTOURA DO NASCIMENTO

DETECÇÃO DE ANOMALIAS EM POÇOS DE PRODUÇÃO DE PETRÓLEO *OFFSHORE* COM A UTILIZAÇÃO DE *AUTOENCODERS* E TÉCNICAS DE RECONHECIMENTO DE PADRÕES

LAVRAS – MG

2021

RODRIGO SCORALICK FONTOURA DO NASCIMENTO

DETECÇÃO DE ANOMALIAS EM POÇOS DE PRODUÇÃO DE PETRÓLEO *OFFSHORE* COM A UTILIZAÇÃO DE *AUTOENCODERS* E TÉCNICAS DE RECONHECIMENTO DE PADRÕES

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração em Sistemas Inteligentes para a obtenção do título de mestre.

Prof. DSc. Bruno Henrique Groenner Barbosa
Orientador

DSc. Ricardo Emanuel Vaz Vargas
Coorientador
DSc. Ismael Humberto Ferreira dos Santos
Coorientador

LAVRAS – MG
2021

Ficha catalográfica elaborada pelo Setor de Repositório Insitucional da Biblioteca Universitária da UFLA

Nascimento, Rodrigo Scoralick Fontoura do.

Detecção de anomalias em poços de produção de petróleo *offshore* com a utilização de *autoencoders* e técnicas de reconhecimento de padrões / Rodrigo Scoralick Fontoura do Nascimento. - 2021.

88 p. : il.

Orientador(a): Bruno Henrique Groenner Barbosa.

Coorientador(a): Ricardo Emanuel Vaz Vargas, Ismael Humberto Ferreira dos Santos.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Reconhecimento de padrões. 2. Detecção de falhas. 3. Validação cruzada. I. Barbosa, Bruno Henrique Groenner. II. Vargas, Ricardo Emanuel Vaz. III. Santos, Ismael Humberto Ferreira dos. IV. Título.

RODRIGO SCORALICK FONTOURA DO NASCIMENTO

DETECÇÃO DE ANOMALIAS EM POÇOS DE PRODUÇÃO DE PETRÓLEO *OFFSHORE* COM A UTILIZAÇÃO DE *AUTOENCODERS* E TÉCNICAS DE RECONHECIMENTO DE PADRÕES

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Engenharia de Sistemas e Automação, área de concentração em Sistemas Inteligentes para a obtenção do título de mestre.

APROVADA em 30 de abril de 2021.

Prof. DSc. Danton Diego Ferreira	UFLA
Prof. DSc. Luiz Henrique de Campos Merschmann	UFLA
Prof. DEng. Murillo Ferreira dos Santos	CEFET-MG

Prof. DSc. Bruno Henrique Groenner Barbosa
Orientador

DSc. Ricardo Emanuel Vaz Vargas
Co-Orientador

DSc. Ismael Humberto Ferreira dos Santos
Co-Orientador

LAVRAS – MG
2021

À minha avó Margarida.

AGRADECIMENTOS

Primeiramente agradeço a Deus por tudo.

À Universidade Federal de Lavras, ao Departamento de Engenharia e ao Programa de Pós-Graduação pela oportunidade de realização do mestrado.

Ao Professor Bruno, pela orientação, pelos ensinamentos e pela perseverança que despertou em mim, para a conclusão deste trabalho.

Ao Ricardo, por todo apoio dado na parte técnica e pelos direcionamentos durante as dificuldades, demonstrando ser um eminente profissional.

Ao Ismael, pelo apoio e disponibilidade ao longo do trabalho.

Agradeço à Petrobras, empresa da qual tenho orgulho de fazer parte porque incentiva a qualificação de seus funcionários.

Aos meus familiares, especialmente à minha esposa Juliene e ao meu filho Miguel, que me acompanharam ao longo desta jornada, incentivando-me.

Por fim, agradeço a todos que contribuíram para o desenvolvimento deste trabalho.

“A persistência é o menor caminho do êxito”.
(Charles Chaplin)

RESUMO

O segmento de Exploração e Produção (E&P) *offshore* da indústria de Óleo & Gás é responsável pela maior parte da produção de petróleo e gás no Brasil. Devido ao elevado nível de complexidade nessa indústria, ela vem demandando novas tecnologias ao longo dos últimos anos. O presente trabalho tem como objetivo o desenvolvimento de sistemas para detecção e classificação de falhas (anomalias) em poços de produção de petróleo *offshore* baseados em *autoencoders* empilhados. Dois conjuntos de dados reais são considerados neste trabalho. O primeiro conjunto é de um poço operado com elevação artificial por *gas lift*, conjunto este obtido durante a execução deste trabalho. Já no segundo, são utilizados os dados de domínio público 3W *dataset* que foram coletados de poços de produção com elevação natural. Em poços com elevação artificial por *gas lift*, são aplicados *autoencoders* empilhados para redução de dimensionalidade e técnicas de reconhecimento de padrões como *k*-NN e Árvore de Decisão para uma falha desconhecida em um poço de petróleo. Após o desenvolvimento destes classificadores, parte dos valores de *recall* obtidos são superiores a 0.98, o que mostra a aplicabilidade do sistema proposto em detectar falhas em poços de produção não surgentes. Para poços surgentes, foram utilizados também os *autoencoders* empilhados para redução de dimensionalidade. Os dados após terem sido tratados foram utilizados como entradas para classificadores de apenas uma classe (*one-class*) como SVM e Floresta de Isolamento de forma a detectar anomalias no processo, como hidrato em linha de produção. Os resultados das médias *F1 score* correspondentes aos modelos utilizados são comparados com outros trabalhos da literatura, nos quais é observada uma melhora em relação às outras abordagens propostas. Os *autoencoders* foram eficazes na redução de dimensão em problemas de detecção e classificação de anomalias em poços de produção de petróleo *offshore* apresentando resultados satisfatórios.

Palavras-chave: *Autoencoders*; Detecção de falhas; Monitoramento de poços de petróleo; Classificação multivariada de séries temporais; Validação cruzada; Reconhecimento de padrões.

ABSTRACT

The Exploration and Production (E&P) offshore segment of the Oil & Gas industry is responsible for most of the oil and gas production in Brazil. Due to the high level of complexity in this industry, it has been demanding new technologies over the past few years. This work aims to develop systems for detecting and classifying failures (anomalies) in offshore oil production wells. Two real data sets are considered in this work. The first set consists of wells operated with artificial lift by gas lift. The second approach uses public domain data 3W dataset that were collected from production wells with natural elevation. Stacked autoencoders are used in artificial elevation gas lift wells to reduce dimensionality and pattern recognition techniques such as k-NN and decision tree for an unknown failure in an oil well. After the development of these classifiers, part of the recall values obtained are greater than 0.98, which shows the applicability of the proposed system in detecting flaws in non-emergent production wells. For emergent wells, stacked autoencoders were used to reduce dimensionality. The data after the treatment were used as inputs for classifiers of only one class (one-class) such as SVM and isolation forest in order to detect anomalies in the process as hydrate in the production line. The results of the F1 score averages presented by the models are compared with other works published in journals and congresses where an improvement is observed in relation to the other proposed approaches. Autoencoders were effective for problems in detecting and classifying anomalies in offshore oil production wells, presenting satisfactory results.

Keywords: Autoencoders; Fault detection; Monitoring of oil wells; Multivariate classification of time series; Cross-validation; Pattern recognition.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

LISTA DE SIGLAS

ANM	Árvore de Natal Molhada
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
APC	*Advanced Process Control*
BSW	*Basic Sediment and Water*
DHSV	*Down Hole Safety Valve*
E&P	Exploração e Produção
IA	Inteligência Artificial
ICA	*Independent Component Analysis*
k-NN	*k-Nearest Neighbor*
MAE	Monitoramento de Alarmes Especialistas
NaN	*Not a Number*
OPEP	Organização dos Países Exportadores de Petróleo
P&ID	*Piping and Instrument Diagram*
PCA	*Principal Component Analysis*
PDG	*Permanent Downhole Gauge*
PI	*Plant Information*
RNA	Redes Neurais Artificiais
SCADA	*Supervisory Control and Data Acquisition*
SDV	*Shutdown Valve*
SIA	Sistema Imunológico Artificial
SVM	*Support Vector Machine*
TPT	Transmissor de Pressão e Temperatura
UFLA	Universidade Federal de Lavras
UN-ES	Unidade de Negócios do Espírito Santo

1 INTRODUÇÃO

O petróleo é uma grande fonte de energia do planeta, utilizado em todas as nações por meio dos seus derivados, sendo insumo para as mais variadas indústrias. Considerado como "ouro negro", o petróleo foi base de uma revolução industrial e motivo para diversas guerras ao redor do mundo.

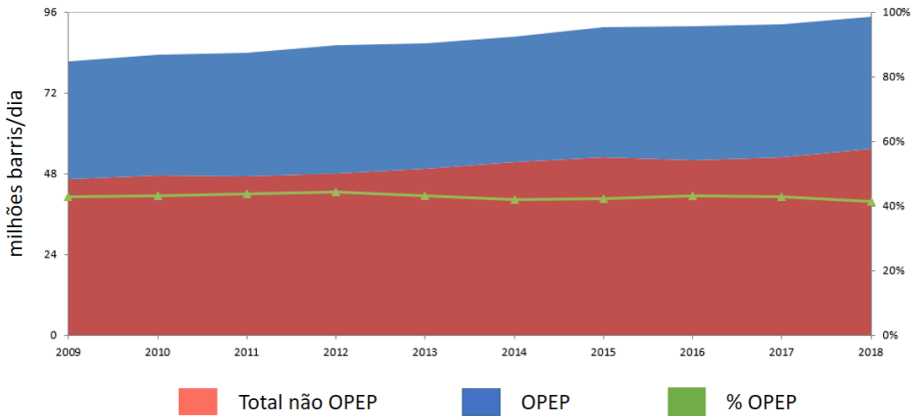
No Brasil, os primeiros poços foram descobertos na década de 1930 no estado da Bahia. Já no ano de 1953 foi criado o monopólio estatal da pesquisa e lavra, refino e transporte dos seus derivados o que deu origem a criação da Petróleo Brasileiro S.A. (Petrobras) (PEYERL, 2017).

Já nas décadas de 1960 e 1970 foram descobertos campos de produção marítimos (LUCCHESI, 1998). Em 2006, a Petrobras anunciou a descoberta do pré-sal, campos gigantescos localizados abaixo da camada de sal no fundo do oceano, que viabilizaram a criação de uma área industrial responsável pela movimentação de bilhões de dólares não somente no Brasil, mas em todo o mundo.

A produção de petróleo no Brasil no mês de fevereiro de 2021 foi de cerca de 2,819 MMbbl/d (milhões de barris por dia) de acordo com a ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis). O petróleo produzido na camada do pré-sal já representa mais de 71% do total produzido no país (ANP, 2021). Os gráficos apresentados nas Figuras 1.1, 1.2, 1.3 e 1.4 mostram o avanço da indústria petrolífera nos cenários nacional e internacional.

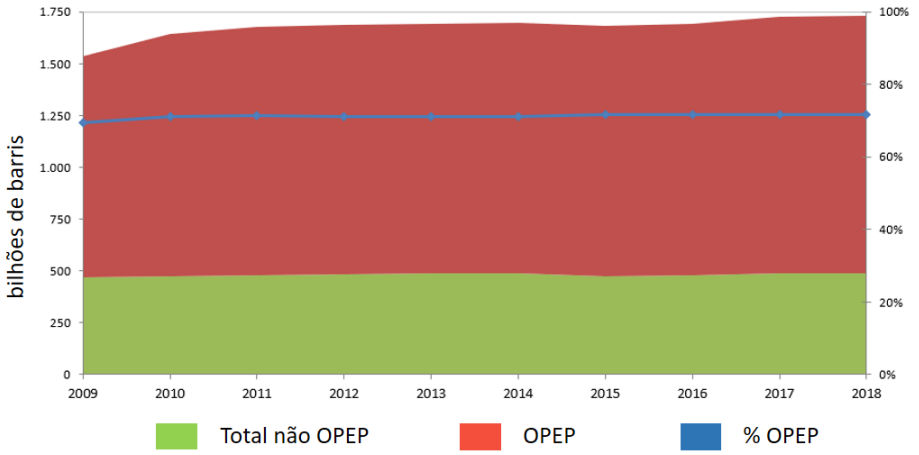
A Figura 1.1 apresenta a produção de petróleo mundial, na qual também está inserida a da OPEP (Organização dos Países Exportadores de Petróleo). Pode-se notar um aumento progressivo da produção ao longo dos anos. Na Figura 1.2 é exibida a evolução das reservas provadas no planeta, onde há um aumento gradual com o tempo. A Figura 1.3 mostra a evolução da produção de petróleo no Brasil e a Figura 1.4 apresenta a evolução das reservas provadas no Brasil, onde nota-se um crescimento na produção de petróleo ao longo dos anos, confirmando o aumento de demanda, tanto nacional quanto internacional. Constata-se, a partir da análise dos gráficos indicados nas figuras, que o crescimento da produção no Brasil deve-se à descoberta do pré-sal a partir da década de 2010.

Figura 1.1 – Evolução da produção de petróleo mundial – 2009-2018.



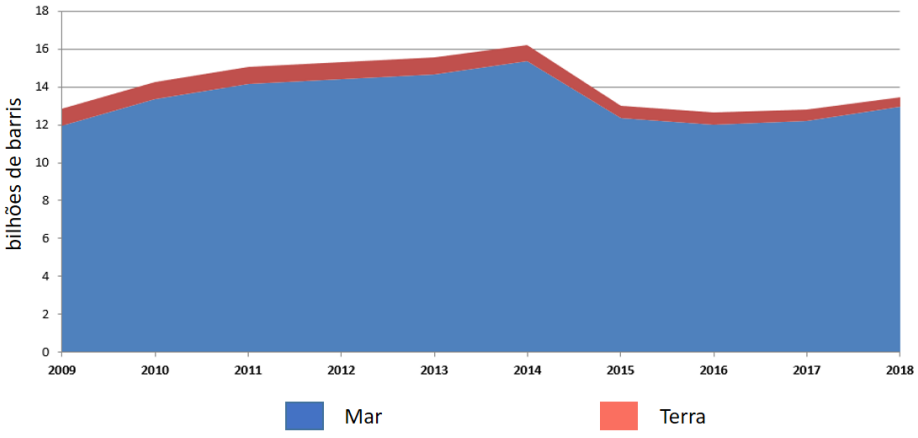
Fonte: (ANP, 2019).

Figura 1.2 – Evolução das reservas provadas de petróleo mundial – 2009-2018.



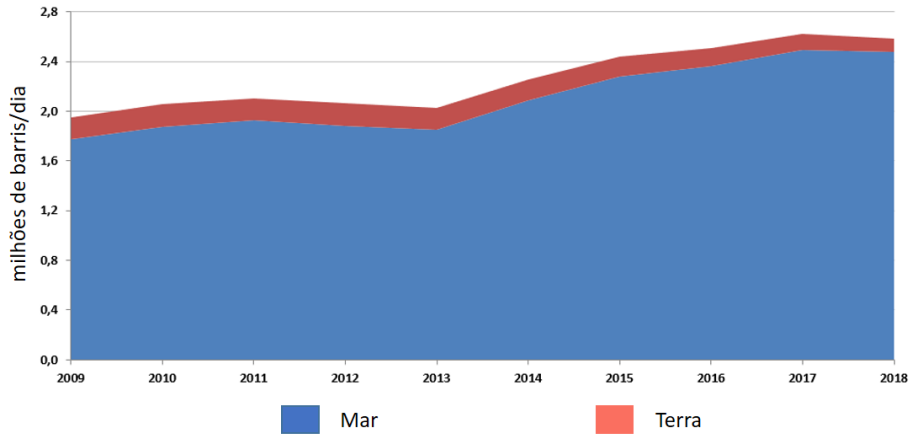
Fonte: (ANP, 2019).

Figura 1.3 – Evolução das reservas de petróleo por localização (terra e mar) no Brasil – 2009-2018.



Fonte: (ANP, 2019).

Figura 1.4 – Evolução da produção de petróleo por localização (terra e mar) no Brasil – 2009-2018.



Fonte: (ANP, 2019).

Os campos de petróleo, inseridos na camada do pré-sal, tendem a gerar um crescimento da produção ao longo dos próximos anos, fortalecendo a economia do país, agregando tecnologia aos processos de construção, montagem e operação, qualificando os envolvidos direta e indiretamente na indústria de óleo e gás.

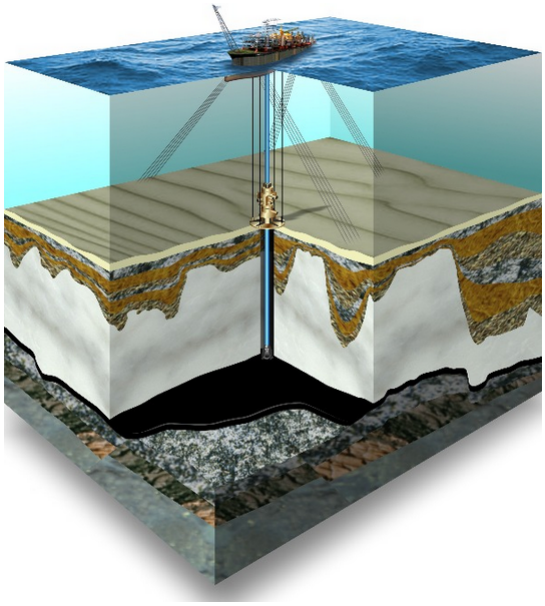
À conta de toda evolução na descoberta de petróleo, no cenário atual, a indústria petrolífera tem se tornado mais exigente em todas as áreas da engenharia, dentre elas as de segurança e de produção. Diversos aspectos devem ser levados em consideração na área de petróleo e gás devido a sua complexidade. Ela engloba vários campos da engenharia, que se relacionam em busca de processos e produtos de melhor qualidade, agregando tecnologia e inovação ao longo dos anos de desenvolvimento das técnicas de extração e uso do produto, como abordado por Schiavi e Hoffmann (2015).

A indústria de extração de petróleo é dividida em duas modalidades de produção: *onshore* e *offshore*. A primeira é baseada na produção em terra firme, no continente. Já a segunda modalidade é realizada no mar por meio de plataformas de extração de petróleo, geralmente distantes do continente e em águas de alta profundidade, sendo essa modalidade tema dissertado neste presente trabalho.

De acordo com a ANP, os campos marítimos foram responsáveis por cerca de 96% da produção de petróleo nacional no mês de fevereiro de 2021 (ANP, 2021). Assim como cresce a produção do insumo petrolífero, cresce também a necessidade de ampliar os avanços tecnológicos em toda a cadeia de projeto e concepção de métodos e equipamentos complexos. Fazendo com que estes equipamentos sejam capazes de alcançar áreas cada vez mais profundas

do pré-sal, podendo atingir mais de sete mil metros de profundidade abaixo do nível do mar, como exibido na Figura 1.5.

Figura 1.5 – Plataforma de produção do pré-sal.



Fonte: Paulo Cabral/Banco de imagens da Petrobras.

A ocorrência de falhas em poços de produção de petróleo *offshore* pode gerar prejuízos de milhares de dólares para empresas produtoras (SANTOS et al., 2018). Além disso, uma complexa operação pode se suceder após a ocorrência das falhas, a fim de que se restabeleça a normalidade na operação dos poços. Uma forma de prever o surgimento destas falhas é a implementação de sistemas de reconhecimento de padrões baseados em Inteligência Artificial (IA).

A IA é determinada como o ramo da Ciência da Computação que se ocupa do comportamento inteligente, conforme discutido por Kaplan e Haenlein (2019). Ou ainda, a IA pode ser comparada com uma inteligência similar à humana, obtida por meio de sistemas computacionais. A IA também pode ser definida conforme Luger (2004) como uma capacidade do sistema para interpretar corretamente dados externos, aprender a partir desses dados e utilizar essas aprendizagens para atingir objetivos e tarefas específicos por meio de adaptação flexível. A área de *Machine Learning* (Aprendizado de Máquina) apresenta-se como uma subárea da IA que trata sobre o aprendizado de sistemas de maneira autônoma com pouca ou nenhuma intervenção humana.

Já o Aprendizado de Máquina é definido em Simon (2013) como o campo de estudo que fornece aos computadores a habilidade de aprender sem serem explicitamente programados. O

aprendizado deve vir de experiências adquiridas por meio de uma base de dados, conseguindo identificar padrões e gerar tomadas de decisões de acordo com o problema a ser resolvido.

1.1 Motivação

Por meio da revolução da informação, o mundo vem se modificando e os dados são gerados de forma abundante. Surge então o conceito de *Big Data*, que descreve o enorme volume de dados, estruturados ou não, que impacta os negócios em qualquer empresa no seu dia a dia (MACHADO, 2018). A informatização da indústria traz benefícios no modo de produzir, sendo a Indústria 4.0 um exemplo de como estas modificações podem desenvolver amplamente o modo de produção. Estes conceitos são fundamentais para a aplicação de técnicas que facilitam a operação na indústria de óleo e gás, mais especificamente em plataformas de produção de petróleo, onde se pretende automatizar os mais diversos processos em busca de melhorias contínuas no negócio de exploração de petróleo *offshore*.

No presente trabalho, o Aprendizado de Máquina se torna uma ferramenta adequada, a qual pode propiciar mecanismos para aplicações de detecção de falhas em poços de produção de petróleo *offshore*, utilizando-se séries temporais. Com a implementação de algoritmos inteligentes, trabalha-se de forma *online* na monitoração da produção, com atuação direta no controle dos mecanismos supervisórios e na operação das plantas de processo.

Séries temporais (EHLERS, 2007) são obtidas a partir da observação das variações nos equipamentos e sistemas supervisórios para auxiliar na operação e diagnóstico de anomalias. Com a análise de especialistas da área de Elevação e Escoamento de petróleo, são obtidas tendências de eventos indesejáveis. A partir destas observações podem ser desenvolvidos detectores de anomalias por meio de técnicas de reconhecimento de padrões . Trata-se de um processo complexo, onde diferentes tipos de falhas podem ocorrer. A escolha dos atributos da base de dados a serem utilizados é bastante crítica. Também é necessário o uso de ferramentas computacionais na construção de modelos que consigam identificar estas anomalias de forma automática. Nesse contexto, os *autoencoders* são técnicas bastante empregadas na detecção de falhas e redução de dimensionalidade em processos industriais, como apresentado em Park et al. (2019), Lu et al. (2016) e Fan, Wang e Zhang (2017).

Como exemplo da aplicação de uma técnica em detecção de eventos indesejáveis na indústria de petróleo e gás, abordado por Araujo, Aguilar e Aponte (2003), um Sistema Imunológico Artificial (SIA) foi desenvolvido para detecção de falhas em poços de produção de

petróleo com elevação artificial por *gas lift* . Com o objetivo da da rotulagem das classes são relacionados alguns tipos de operações da produção (normais e anormais) das pressões nos poços, utilizando 80% dos dados para geração dos detectores de falhas e 20% para testes do sistema. Este trabalho gerou detectores que determinam desvios no processo de produção.

Já Santos et al. (2018) propõem a detecção de eventos indesejáveis de acumulação de hidrato em linhas de injeção de água ou produção de petróleo *offshore*. Anomalias que trazem a diminuição de vazão de óleo ou até perda total da produção do poço. Detectam-se as falhas em poços de produção surgentes utilizando duas abordagens. A primeira baseada em dados, e a segunda baseada em modelos, entretanto ambas trabalhando em paralelo. A abordagem baseada em dados separa as classes em falhas normais ou falhas por hidrato, implementando a técnica de Floresta Aleatória e verificando o desempenho com a métrica *F1 score*, utilizando validação cruzada por meio do método *k-fold* e usando aprendizado supervisionado. Na abordagem baseada em modelo foi utilizado o método de Análise do Componentes Principais. Para a eliminação de ruídos e redundância, foi aplicado um índice de detecção de falhas como filtro. Com o objetivo de se retirar alarmes correlacionados, esse trabalho foi capaz de diferenciar o comportamento normal e o defeituoso com 77% de acurácia. As 12 falhas por hidrato analisadas foram detectadas no trabalho, sendo que 85% delas de forma antecipada.

O presente trabalho tem por finalidade a implementação de sistemas de reconhecimento de padrões baseados em técnicas de Inteligência Computacional, tornando o processo mais analítico e menos operacional, sendo este uns dos objetivos principais da Indústria 4.0 por meio de tecnologias disruptivas (NILCHIANI; EDWARDS; GANGULY, 2019).

1.2 Objetivos

1.2.1 Objetivo Geral

A ideia central da dissertação é a identificação de eventos indesejáveis em poços de produção com elevação artificial por *gas lift* e poços surgentes com uso de *autoencoders*. Para poços surgentes é proposta uma nova abordagem para o segundo *benchmark* proposto por Vargas et al. (2019).

1.2.2 Objetivos Específicos

Os objetivos específicos são:

- Verificar a aplicabilidade de *autoencoders* para redução de dimensionalidade em séries temporais de variáveis de poços de petróleo;
- Realizar comparações de modelos obtidos com outros já desenvolvidos, referenciando o seu desempenho;
- Verificar a influência do número de entradas reduzidas (atributos) no desempenho final dos classificadores e detectores;
- Analisar a influência do tipo de falha e suas peculiaridades no desempenho geral do detector de anomalias.

2 REFERENCIAL TEÓRICO

Este capítulo consiste na realização de uma revisão bibliográfica dos temas relacionados aos processos de produção de petróleo abordados e ferramentas de inteligência computacional, reconhecimento de padrões, análise e desempenho que são utilizadas ao longo do trabalho.

2.1 O Processo de Elevação Petróleo *Offshore*

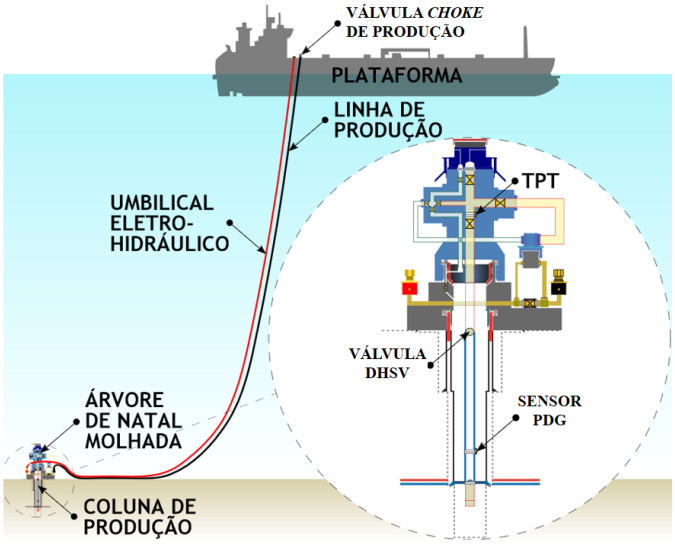
Elevação é o termo utilizado na indústria petrolífera para caracterizar o processo de ascensão do fluido contido em um reservatório até a superfície (THOMAS, 2004). Os poços de petróleo em águas profundas classificam-se em surgentes e não surgentes. Poços não surgentes necessitam de métodos para auxiliar o escoamento dos fluidos (água, óleo, gás e sedimentos); enquanto que os poços surgentes conseguem, com sua própria pressão, realizar o escoamento dos fluidos de produção, ou seja, nos poços surgentes há uma elevação natural dos fluidos (THOMAS, 2004).

Para aumentar a produção de petróleo e a viabilidade da produção nos poços de petróleo, são utilizados determinados métodos de elevação artificial. Esses métodos têm por finalidade auxiliar no escoamento da produção dos poços de petróleo. A técnica de elevação artificial é largamente utilizada na indústria *offshore* nacional (FILHO, 2011).

Thomas (2004) diz que a seleção do melhor método de elevação artificial para um determinado poço ou campo depende de vários fatores. Os principais a serem considerados são: número de poços, diâmetro do revestimento, produção de areia, razão gás-líquido, vazão, profundidade do reservatório, viscosidade dos fluidos, mecanismo de produção do reservatório, disponibilidade de energia, acesso aos poços, distâncias dos poços às estações ou plataforma de produção, equipamento disponível, pessoal treinado, investimento, custo operacional, segurança, entre outros. Com estas informações é possível avaliar e indicar a melhor técnica para utilização no reservatório.

A Figura 2.1 reproduz um esquema simplificado de um poço de produção de petróleo com elevação natural dos fluidos. Nos poços não surgentes há uma baixa pressão no reservatório, tal que o escoamento dos fluidos não é possível, caso não haja um sistema de elevação artificial para auxiliar na produção. Ocorre que, ao longo do tempo, poços surgentes podem necessitar de métodos de elevação artificial caso sejam economicamente viáveis, pois tendem a perder a pressão no fundo do poço. Assim sendo, o método de elevação artificial por *gas lift* é a forma mais utilizada na indústria nacional (FILHO, 2011).

Figura 2.1 – Esquema simplificado de um poço surgente *offshore*.



Fonte: Adaptado de (VARGAS et al., 2019).

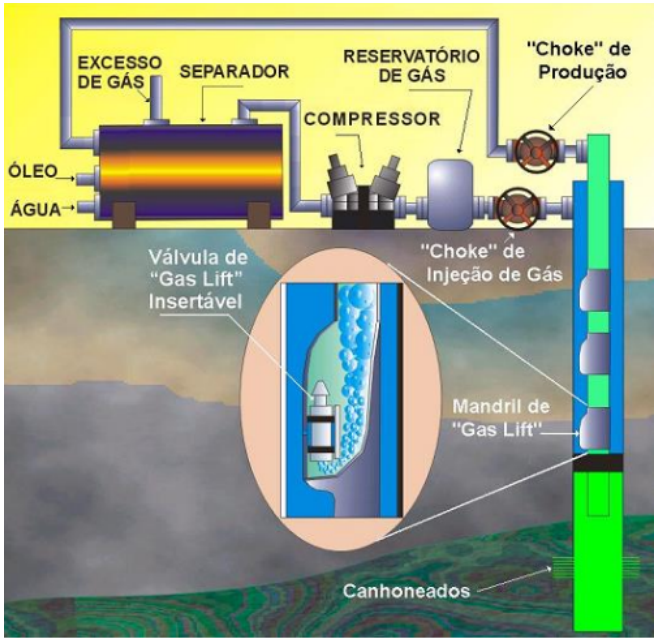
O processo de elevação artificial por *gas lift* consiste basicamente na injeção de gás com uma pressão elevada, tendo como finalidade diminuir a densidade média dos fluidos produzidos aumentando a vazão na linha de produção. Trata-se de um processo complexo e sujeito a diversas falhas, sejam elas dos atuadores, sensores ou de acordo com as características físico-químicas dos poços de produção.

Filho (2011) destaca que os poços que empregam a tecnologia do *gas lift* não eram a maioria em 2005 e representavam apenas 2% do número de poços do Brasil. Devido às suas vantagens frente às demais técnicas de elevação artificial e com a descoberta de novos campos de extração, em 2009, a tecnologia representava mais de 70% da produção em termos de petróleo produzido no país (FILHO, 2011). O *gas lift* é um método de elevação artificial que utiliza a energia contida em gás comprimido para elevar fluidos (óleo, água, gás e sedimentos) até a superfície. Plucenio (2003) explica que nesse método, a força motriz da produção é o diferencial de pressão entre o fundo da coluna de produção e a pressão de operação do separador. Sendo injetado o gás por meio da válvula de *gas lift*, as válvulas *choke* controlam os fluxos do fluidos, conforme a Figura 2.2.

O gás é utilizado para gaseificar a coluna de fluido (*gas lift* contínuo) ou simplesmente para deslocá-la (*gas lift* intermitente) de uma determinada profundidade até a superfície. É um método muito versátil em termos de vazão (1 a 1700 m^3/h) e profundidade (2600 m, dependendo da pressão do gás de injeção) sendo propício para poços que produzem fluidos com alto teor de

areia, elevada razão de gás-líquido e por exigir investimentos relativamente baixos para poços profundos (THOMAS, 2004).

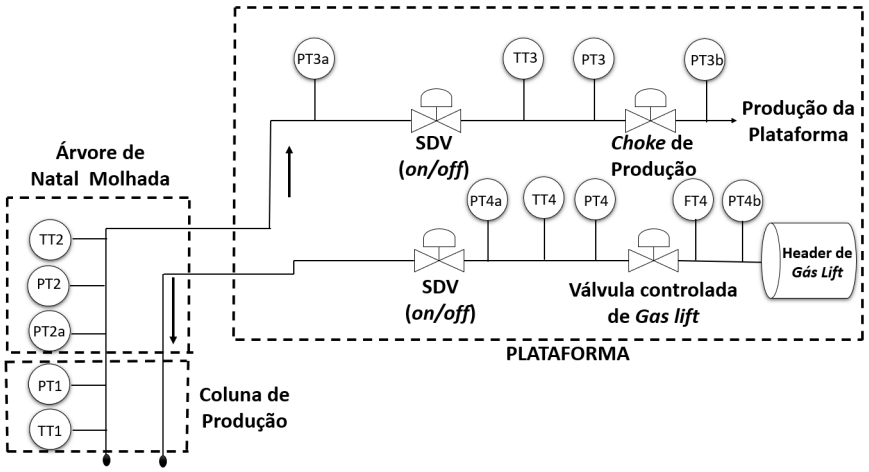
Figura 2.2 – Sistema de *gas lift* contínuo simplificado.



Fonte: (PLUCENIO, 2003).

A Figura 2.3 mostra um esquema simplificado do diagrama de tubulação e instrumentação P&ID (do inglês, *Piping and Instrument Diagram*) de um poço de produção que utiliza elevação artificial por *gas lift*. O P&ID é um diagrama utilizado em processos industriais, no qual uma das funções é mostrar os instrumentos e equipamentos instalados na planta industrial. Na Tabela 2.1 são apresentadas as variáveis presentes e suas unidades de grandeza.

Figura 2.3 – Diagrama P&ID simplificado de um poço de produção com elevação artificial por *gas lift*.



Fonte: próprio autor.

Aguirre et al. (2017) apresenta o funcionamento do sistema, sendo o gás de alta pressão proveniente do *header* de gás na plataforma, instrumentos marcados por 4 na Figura 2.3. O gás é injetado por meio do anel entre a tubulação e a cadeia de revestimento até atingir uma válvula de orifício localizada na coluna de produção na parte inferior de sua tubulação.

A densidade do fluido é reduzida, o que causa redução do gradiente de pressão médio ao longo da coluna e reduz a energia (pressão) necessária para que os fluidos do reservatório cheguem à plataforma. No leito do oceano, um conjunto de válvulas conhecido como Árvore de Natal Molhada (ANM) permite a passagem dos fluidos para a plataforma e atua como barreira de segurança. Já na plataforma, a válvula de desligamento SDV (*Shutdown Valve*) é instalada para interromper a produção durante uma situação de emergência e a válvula *choke*, de estrangulamento, regula a taxa de fluxo de produção. Diferentes dinâmicas de fluxo são obtidas a depender dos valores das pressões de elevação de gás (PT4a e PT4b) e de fundo de poço (PT1).

Tabela 2.1 – Variáveis do processo para obter modelos de poços com elevação de artificial por *gas lift* . Estes *tags* podem ser vistos na Figura 2.3.

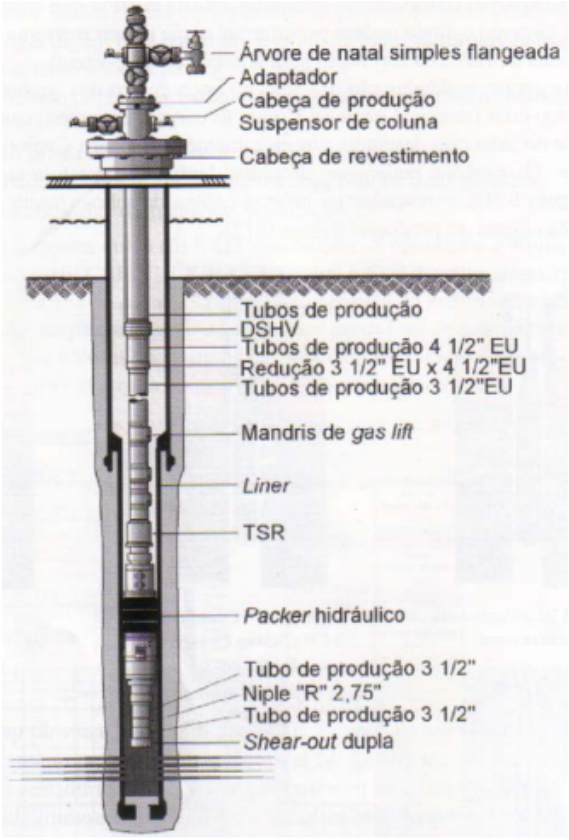
Tag	Descrição	Unidades
PT1	Pressão PDG	kgf/cm^2
TT1	Temperatura PDG	°C
PT2	Pressão na Árvore de Natal Molhada	kgf/cm^2
TT2	Temperatura na Árvore de Natal Molhada	°C
PT2a	Pressão anular	kgf/cm^2
PT3a	Pressão a montante da SVD de Produção	kgf/cm^2
PT3	Pressão a montante da válvula *Choke* de Produção	kgf/cm^2
TT3	Temperatura a montante da válvula *Choke* de Produção	°C
PT4a	Pressão a montante da SVD de *gas lift*	kgf/cm^2
TT4	Temperatura a montante da SVD de *gas lift*	°C
FT4	Fluxo de *gas lift* instantâneo	m^3/h
FV4	Posição da válvula de *gas lift*	%
PT4	Pressão a jusante da válvula de *gas lift*	kgf/cm^2
PT4b	Pressão do *header* de *gas lift*	kgf/cm^2
SDVL	Atuação da SDV de *gas lift*	*on/off*
SDVP	Atuação da SDV de produção	*on/off*

Fonte: próprio autor.

A ANM é um conjunto submarino composto com várias válvulas operadas remotamente por meio de comandos hidráulicos, onde estão contidos por exemplo os sensores TPT (Transmissor de Pressão e Temperatura, TT2) e o sensor que mede a pressão anular na válvula (PT2a) de *gas lift*. As pressões e temperaturas também são aferidas de acordo com a necessidade, bem como a montante e jusante das SDVs, válvula *choke* e válvula de injeção de *gas lift*.

A válvula *choke* é um instrumento de controle de fluxo a jusante. Já a válvula de *gas lift* é um dispositivo destinado a auxiliar no controle da vazão de gás do anular para a coluna do poço. As altas pressões de *gas lift* vêm dos compressores da própria instalação. Tais equipamentos citados neste parágrafo estão instalados na plataforma de produção *offshore*. Uma descrição mais detalhada sobre este processo é feita por (AGUIRRE et al., 2017).

Figura 2.4 – Coluna de produção convencional utilizando a técnica de elevação artificial por *gas lift*.

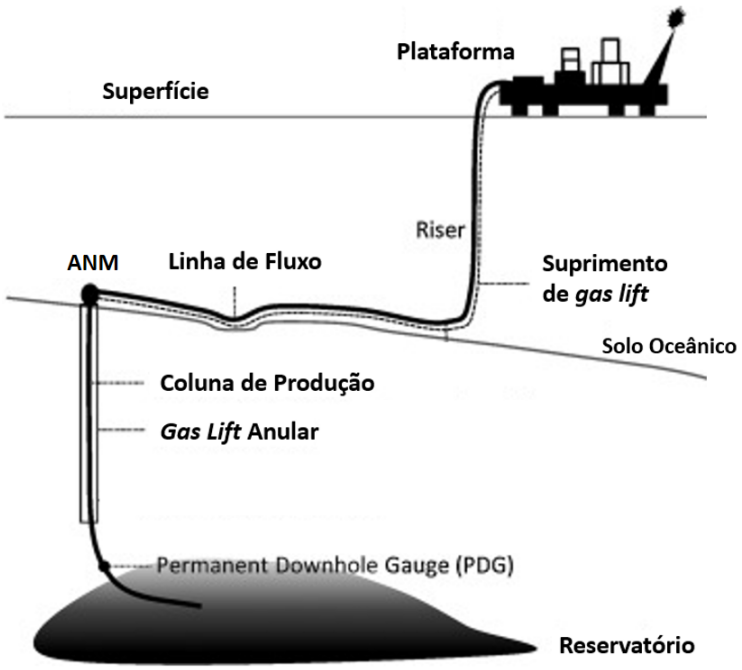


Fonte: (THOMAS, 2004).

A Figura 2.4 exibe uma coluna de produção convencional, onde é utilizada a técnica de elevação artificial por *gas lift*. A coluna de produção é constituída basicamente por tubos metálicos, aos quais são conectados os demais componentes. Ela é inserida no interior do revestimento de produção com as seguintes finalidades básicas: conduzir os fluidos produzidos até a superfície, permitir a instalação de equipamentos para elevação artificial e possibilitar a circulação de fluidos para amortecimento do poço (THOMAS, 2004). Na coluna de produção estão instalados os sensores PDG conforme apresentado na Figura 2.5.

As pressões e temperaturas também são aferidas de acordo com a necessidade, como as pressões e temperaturas a montante das SDVs, válvula *Choke* e válvula de injeção de *gas lift*.

Figura 2.5 – Localização do sensor PDG.



Fonte: Adaptado de (DIEHL et al., 2018).

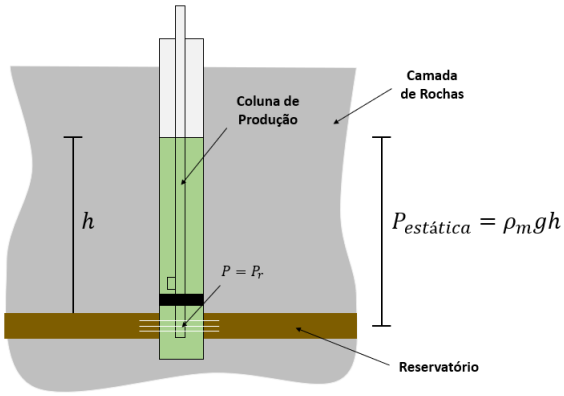
2.2 Falhas em Poços de Produção de Petróleo *Offshore*

Podem ocorrer diversos tipos de falhas em poços de produção de petróleo *offshore*. Essas falhas podem acarretar na perda de produção de fluidos do poço como relatado por Plucenio (2003), onde a pressão estática do reservatório é menor que a pressão da coluna de fluido de formação dentro do tubo de produção. Entre os operadores de produção esta falha é diagnosticada como morte do poço.

O poço não irá produzir se o nível do fluído de formação dentro do tubo de produção estiver em um nível abaixo da parte superior do poço conforme Figura 2.6. Também pode ocorrer uma interrupção nas tubulações de escoamento envolvidas no processo em função de hidratos e incrustações ou defeito em equipamentos que fazem parte da planta do processo de extração de óleo.

A perda de produção ou até a parada do poço *offshore* pode ser ocasionada por diversos fatores que acarretam muitos transtornos para a empresa que extrai o petróleo, os quais geram custos elevados, na casa das centenas de milhares de dólares a diária. Estes prejuízos financeiros são suscitados pela manutenção dos poços, podendo ocorrer desde operações mais simples da própria mão de obra da empresa a até operações mais complexas como barcos especializados nesse tipo de manutenção.

Figura 2.6 – Poço Morto.



Fonte: Adaptado de (PLUCENIO, 2003).

Dentre os fatores geradores de custo estão as incrustações, que podem ser definidas como compostos químicos de natureza inorgânica que se encontram inicialmente dispersos em soluções salinas. Quando essas soluções são submetidas a mudanças de pressão, temperatura e concentração, mudam suas condições de solubilidade (DEMADIS et al., 2007).

A formação de depósitos orgânicos em tubulações é consequência direta da composição dos fluidos produzidos e das condições de temperatura e pressão do escoamento (BRASIL, 2011). Com sistemas de escoamento submetidos a baixas temperaturas por longas distâncias, o resfriamento dos fluidos produzidos pode conduzir por exemplo à precipitação e deposição de parafinas, que na indústria do petróleo são denominadas de forma mais genérica, representando o depósito formado por asfaltenos, sulfetos, água, areia, sais e resinas (OLIVEIRA; GONZALEZ; SANTIAGO, 1998). Normalmente são utilizados dispersantes químicos ou solventes para remover depósitos ou incrustações na tubulação de escoamento.

2.2.1 Tipos de Eventos Indesejáveis em Poços de Produção de Petróleo

Falhas ou anomalias podem ocasionar desde pequenas instabilidades nas linhas de produção à total parada de fluxo do reservatório para a plataforma de produção. A identificação automática dessas falhas pode auxiliar na operação de modo a minimizar perdas em poços de produção de petróleo. Por conseguinte, diminuindo custos de manutenção, encurtando tempo de atuação no problema, evitando gastos adicionais e operações complexas no retorno à operação normal dos sistemas de produção. Alguns tipo de falhas são apresentadas a seguir, como abordado em Vargas et al. (2019).

1. Aumento abrupto de *BSW*: O Sedimento Básico e Água (do inglês *BSW, Basic Sediment and Water*) é definido pela porcentagem de água e sedimentos em relação ao total de fluidos produzidos. A fim de diminuir a perda de produção de petróleo ao longo dos anos de vida de produção de poços de petróleo, é injetada água em alta pressão. Essa água injetada no aquífero do reservatório natural tende a aumentar o *BSW* no ciclo de produção do poço. De acordo com Vargas et al. (2019) o aumento abrupto do *BSW* pode acarretar vários problemas relacionados à garantia de fluxo como menor produção de petróleo, escoamento de óleo, incrustação e fator de recuperação. A identificação deste evento indesejável de forma automática permite ações na atuação da operação da produção de fluidos e injeção de água de forma a minimizar ou até evitar este tipo de falha;

2. Fechamento espúrio de *DHSV*: Válvula de Segurança de Sub Superfície (do inglês, *Down Hole Safety Valve*) é um equipamento localizado na coluna de produção. É uma válvula de segurança que tem o objetivo de evitar erupções ou fluxos descontrolados do poço no caso de alguma falha dos equipamentos de segurança de superfície. Trata-se de uma válvula que opera normalmente aberta, fechando em ocasiões de segurança operacional. Pode ocorrer de forma espúria a atuação dessa válvula, algumas vezes sem indicação no supervisório da plataforma. A identificação dessa falha de modo automático remete modos de correções como a reabertura da válvula por meio da operação do sistema;

3. Intermitência Severa: *Slug* são bolsões de líquido que escoam alternadamente com grandes bolhas de gás, tendo um padrão de escoamento altamente intermitente em frequências aleatórias flutuantes (CARNEIRO et al., 2010). O padrão *slug* induzido no processo é chamado de Intermitência Severa (do inglês, *Severe Slugging*) e acontece quando uma tubulação ligeiramente inclinada encontra uma tubulação vertical (TAITEL; BARNEA, 1990), representando um evento de grande criticidade. As duas características mais marcantes deste evento são as bem definidas periodicidades (em torno de 30, 45 ou 60 minutos) e a intensidade. A intensidade do *Slug* é geralmente suficientemente alta para ser detectada por sensores ao longo de toda a linha de produção (MEGLIO et al., 2012). Dependendo da periodicidade e intensidade, esse tipo de evento pode resultar em estresse ou até mesmo danificar equipamentos no poço e/ou na planta industrial (VARGAS et al., 2019);

4. Instabilidade de Fluxo: Durante uma instabilidade de fluxo, pelo menos uma das variáveis monitoradas sofre mudanças relevantes, mas com amplitudes toleráveis (VARGAS et al., 2019). Uma característica que diferencia este tipo de evento indesejável da Intermitência Severa é a falta de periodicidade entre essas mudanças (THEYAB, 2018). Esta instabilidade pode acarretar a ocorrência do *Severe Slugging*, mas com a identificação da instabilidade evita-se o aumento da criticidade do problema;

5. Perda Rápida de Produtividade: Vários fatores podem influenciar o escoamento de poços de produção surgentes e não surgentes como: reservatório de pressão estática, percentagem de *BSW*, viscosidade do fluido produzido, diâmetro da linha produção, dentre outros (HAUSLER; KRISHNAMURTHY; SHERAR, 2015). Quando essas propriedades são alteradas, até o momento em que a energia do sistema não seja mais suficiente para superar as perdas, o fluxo diminui ou até é interrompido (VARGAS et al., 2019);

6. Restrição Rápida na Válvula *Choke* de Produção: A *Choke* de Produção é uma válvula de controle de fluxo de escoamento dos fluidos produzidos instalada na chegada das linhas de produção na plataforma. Estas válvulas podem operar de forma manual. Neste tipo de operação podem ocorrer restrições rápidas e indesejadas de fluxo, conta de problemas operacionais;

7. Incrustação na Válvula *Choke* de Produção: A válvula *Choke* de Produção é suscetível a incrustações de depósitos inorgânicos. Essa falha depende das características físico-químicas dos poços que podem ser advindas da variação de pH ou temperatura, o que afeta a produção do poço;

8. Hidrato em Linha de Produção: Os hidratos são estruturas cristalinas com aparência de um cristal de gelo, nos quais há dois ou mais componentes associados sem ligações químicas covalentes. Essa associação ocorre por meio de um completo encapsulamento de um tipo de molécula por outra que se origina por meio da junção da água com gases de baixo peso molecular ou hidrocarbonetos de cadeias curtas (SILVA, 2018). O acúmulo de hidrato pode acarretar na interrupção do fluxo de fluidos em linhas de escoamento da produção. A detecção desta falha de modo automático, diminuiria de maneira significativa os custos com manutenção no reestabelecimento da operação normal dos poços.

As anomalias citadas são as falhas mais corriqueiras na indústria petrolífera *offshore*, podendo ter suas ocorrências tanto em poços com elevação natural quanto em poços que uti-

lizam elevação artificial. Estes eventos são característicos dos dois métodos de elevação, pois todos os equipamentos e processos onde acontecem estas anomalias são comuns nos dois métodos de escoamento de fluidos. Vargas et al. (2019) apresenta as estimativas de tamanhos de janelas temporais utilizados para confirmar ocorrências de anomalias, presentes na Tabela 2.2.

Tabela 2.2 – Estimativas de tamanhos de janelas temporais utilizados para confirmar ocorrências de anomalias.

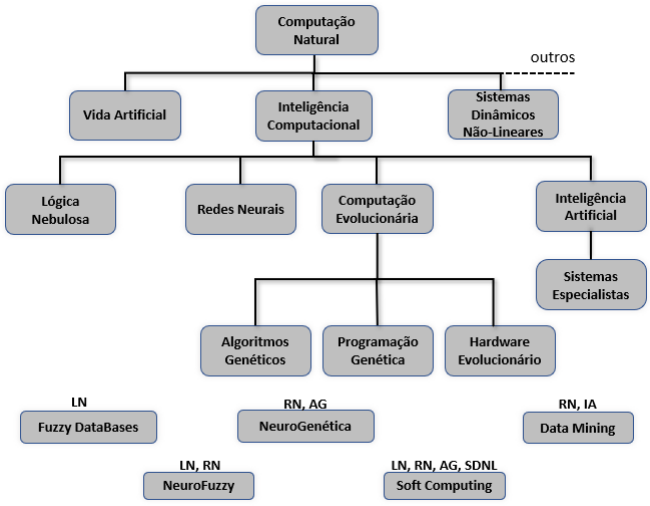
Falha	Anomalia	Tempo de falha
Falha 1	Aumento abrupto de *BSW*	12 h
Falha 2	Fechamento espúrio de *DHSV*	5 min -20 min
Falha 3	Intermitência Severa	5 h
Falha 4	Instabilidade de Fluxo	15 min
Falha 5	Perda Rápida de Produtividade	12 h
Falha 6	Restrição Rápida na Válvula *Choke* de Produção	15 min
Falha 7	Incrustação na Válvula *Choke* de Produção	72 h
Falha 8	Hidrato em Linha de Produção	30 min - 5 h

Fonte: Adaptado de (VARGAS, 2019).

2.3 Inteligência Computacional

A Inteligência Computacional está cada vez mais difundida na sociedade, seja por meio de aplicativos de dispositivos móveis, indústria, internet e outras áreas relacionadas à Ciência da Computação. Segundo Liden (2008), a inteligência computacional procura desenvolver comportamentos similares a certos aspectos do comportamento inteligente.

Figura 2.7 – Inteligência Computacional, uma sub-área da Computação Natural.



Fonte: Adaptado de (CASTRO; ZUBEN, 2005)

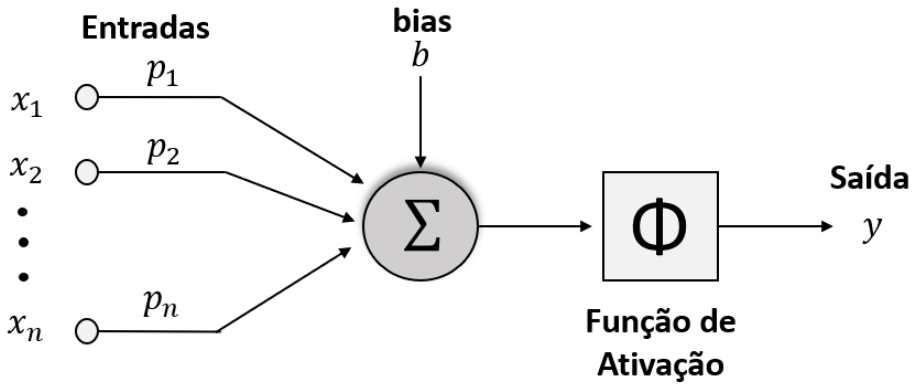
Castro e Zuben (2005) apresentam um organograma no qual está inserida a Inteligência Computacional e suas sub-áreas. Ela é uma grande área da computação natural como é exibido na Figura 2.7, sendo a IA uma sub-área que engloba sistemas especialistas, onde se encontram as ferramentas para detecção e classificação de anomalias por reconhecimento de padrões.

2.3.1 Redes Neurais Artificiais

Uma ferramenta para resolução de problemas e auxílio no reconhecimento de padrões são as Redes Neurais Artificiais (RNAs). Estas redes utilizam neurônios em cadeia para resoluções de problemas. A habilidade de um ser humano em realizar funções complexas e principalmente a sua capacidade de aprender, advêm do processamento paralelo e distribuído da rede de neurônios do cérebro. Os neurônios do córtex, camada externa do cérebro, são responsáveis pelo processamento cognitivo.

Um novo conhecimento ou uma nova experiência pode levar a alterações estruturais no cérebro. Tais alterações são efetivadas por meio de um rearranjo das redes de neurônios, reforçando ou inibindo algumas sinapses (HAYKIN, 2001). O modelo de neurônio artificial da Figura 2.8 é uma simplificação do perceptron, um tipo de rede neural artificial apresentada em 1957 por Frank Rosenblatt .

Figura 2.8 – Modelo matemático de um neurônio.



Fonte: próprio autor.

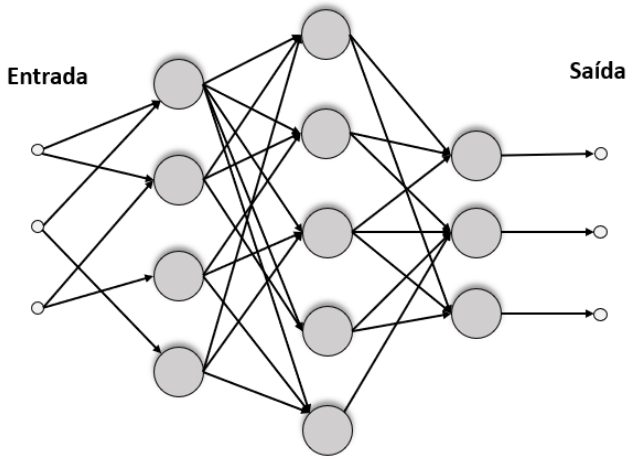
Este modelo é composto por quatro elementos básicos:

- Um conjunto de n conexões de entrada (x_1, x_2, ..., x_n), caracterizadas por pesos (p_1, p_2, ..., p_n);

- O bias b que é utilizado para melhorar o grau de liberdade e aperfeiçoar a adaptação da rede;
- Um somador (Σ) para acumular os sinais de entrada;
- Uma função de ativação (ϕ) que limita o intervalo permissível de amplitude do sinal de saída (y) a um valor fixo.

De acordo com os valores da entrada, pesos e *bias*, o neurônio artificial tem uma saída que pode ser interpretada de várias formas, podendo solucionar vários problemas, dentre eles os de classificação. Para tal, utiliza-se uma rede neural artificial com a combinação de diversos neurônios, formando-se a rede neural artificial que se traduz em modelos que buscam simular o processamento de informação do cérebro humano. As RNAs são compostas por unidades de processamentos simples dos neurônios, que se unem por meio de conexões sinápticas, como exibido na Figura 2.9.

Figura 2.9 – Representação simplificada de uma rede neural artificial.



Fonte: própio autor.

As redes neurais artificiais se diferenciam pela sua arquitetura e pela forma como os pesos associados às conexões são ajustados durante o processo de aprendizado. A arquitetura de uma rede neural restringe o tipo de problema no qual a rede poderá ser utilizada e é definida pelo número de camadas (camada única ou múltiplas camadas), pelo número de nós em cada camada, pelo tipo de conexão entre os nós (*feedforward* ou recorrente) e por sua topologia (HAYKIN, 2001).

Uma das propriedades mais importantes de uma rede neural artificial é a capacidade de aprender por intermédio de exemplos e fazer inferências sobre o que aprendeu, melhorando gradativamente o seu desempenho. As redes neurais utilizam um algoritmo de aprendizagem, como o *backpropagation* (CHAUVIN; RUMELHART, 1995) e o Levenberg-Marquardt (MORÉ, 1978), cuja tarefa é ajustar os pesos de suas conexões (BRAGA; CARVALHO; LUDERMIR, 2007).

2.3.1.1 Funções de Ativação

Uma função de ativação é responsável por limitar o valor de saída do neurônio, geralmente no intervalo [0,1] ou $[-1,1]$. Estas funções são utilizadas como uma função de transferência de valores entre neurônios (SHARMA; RAI; DEV, 2012). Existem diversos tipos de funções de ativação, sendo utilizadas nas mais variadas aplicações, cada uma com sua particularidade. As funções mais recentes desenvolvidas como ReLU (*Rectified Linear Unit*), LeakyReLU, PReLU são apresentadas nos próximos sub-tópicos; além das mais comuns como sigmoidal e tangente hiperbólica que por serem bastante conhecidas não são aqui descritas.

2.3.1.1.1 ReLU

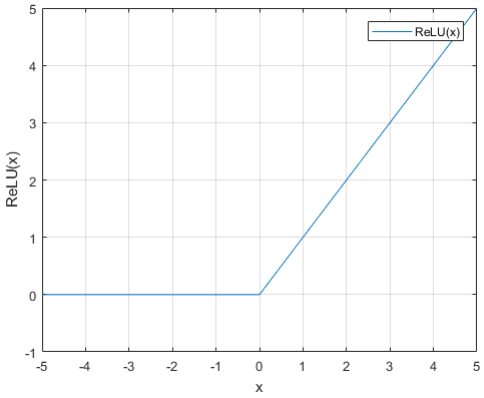
A ReLU, como o próprio nome indica, é uma unidade linear retificada, sendo uma função com limiar em zero caracterizada como não-linear, sendo definida por:

$$f(x) = max(0,x) \tag{2.1}$$

$$f(x) = \begin{cases} 0 & for\, x < 0 \\ x & for\, x \geq 0 \end{cases} \tag{2.2}$$

Este tipo de função de ativação tem as características diferentes em relação a outras funções mais implantadas aos longo dos anos que é de acelerar a convergência do gradiente estocástico (*Stochastic Gradiente Descent*, SGD). Ela pode ser implementada com menor custo computacional. Isso se deve ao fato de não utilizar funções mais complexas como funções exponenciais e ser capaz de inativar funções e inativar neurônios quando a soma ponderada de todas suas entradas for negativa, não propagando assim o erro durante a fase de *backpropagation*.

Figura 2.10 – Gráfico ReLU.



Fonte: próprio autor.

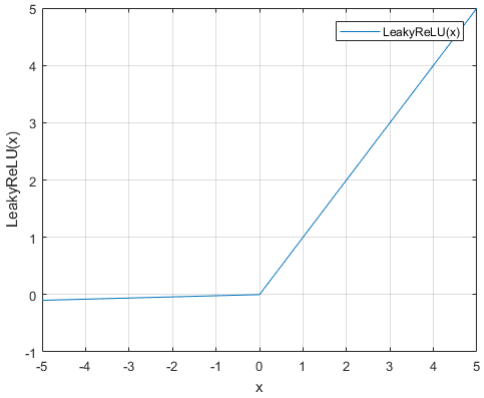
2.3.1.1.2 *Leaky*ReLU

Esta função de ativação é similar a ReLU, na qual até $x < 0$ há uma curva com um coeficiente angular muito pequeno chamado de α . A função de ativação *Leaky*ReLU é dada por:

$$f(x) = \begin{cases} \alpha x & for\ x < 0 \\ x & for\ x \geq 0 \end{cases} \tag{2.3}$$

O gráfico desta função é apresentado com $\alpha = 0.01$, exibido na Figura 2.2. Este tipo de função de ativação possui características similares em relação a ReLU, obtendo um menor custo computacional na sua implementação e sendo capaz de inativar funções e neurônios quando a soma ponderada de todas suas entradas for negativa.

Figura 2.11 – Gráfico LeakyReLU.



Fonte: próprio autor.

2.3.1.1.3 *Parametric* ReLU

A função de ativação *Parametric* ReLU (PReLU) é muito parecida com a função *Leaky*ReLU e tem uma generalização maior que é tornar o coeficiente α em um parâmetro que é aprendido junto com os outros parâmetros da rede neural (HE et al., 2015), sua função segue:

$$f(x) = \begin{cases} \alpha x & for\ x < 0 \\ x & for\ x \geq 0 \end{cases} \tag{2.4}$$

2.3.1.2 *Autoencoder*

Shin et al. (2013) explicam que o *autoencoder* é um tipo de RNA que é formada por três camadas, sendo o *encoder* constituído pelos neurônios das duas primeiras camadas e os neurônios das duas últimas configurando o *decoder* como apresentado na Figura 2.12. Corroborando Yu e Zhang (2020) dizem que o *autoencoder* tem a função de mapear o mais próximo possível a entrada em sua camada de saída. Geralmente os *autoencoders* têm em sua camada oculta um número inferior de neurônios comparado aos das suas camadas de entrada e saída. Isso é benéfico em relação à diminuição da dimensionalidade dos dados, que faz com que o *autoencoder* utilize apenas as principais características dos dados de entrada com o intuito de eliminar descritores de pouca relevância para os modelos. Além de reduzir a dimensão o *autoencoder* também transforma os dados não-linearmente, propiciando a maximização das diferenças entre as classes.

Métodos típicos de análise de dados como o de análise dos componentes principais PCA (do inglês, *Principal Component Analysis*) e análise dos componentes independentes ICA (do inglês, *Independent Component Analysis*) se distinguem dos *autoencoders* por não se desempenharem bem em dados não-lineares. Embora esses métodos baseados em projeção de dados capturem a formação global dos dados, a estrutura local geralmente é ignorada (YU; ZHANG, 2020). Além disso, esses métodos não são eficazes para processos não-lineares, porque os recursos extraídos não são muito eficazes para descrever a distribuição dos sinais de processos complexos.

A Figura 2.12 apresenta a estrutura de um *autoencoder*, cujos dados de entrada são $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]$, os valores de saída do *autoencoder* são $\mathbf{z} = [z_1\ z_2\ \ldots\ z_n]$, sendo n o número de neurônios tanto na camada de entrada quanto na de saída. O vetor $\mathbf{h} = [h_1\ h_2\ \ldots\ h_m]$ é a representação da entrada $\mathbf{x}$ na camada oculta após a utilização de uma função de ativação

sigmóide (sf) e m é a quantidade de neurônios na camada escondida. As equações que regem esse tipo de modelo são descritas por:

$$\mathbf{h} = sf\left(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}\right) \tag{2.5}$$

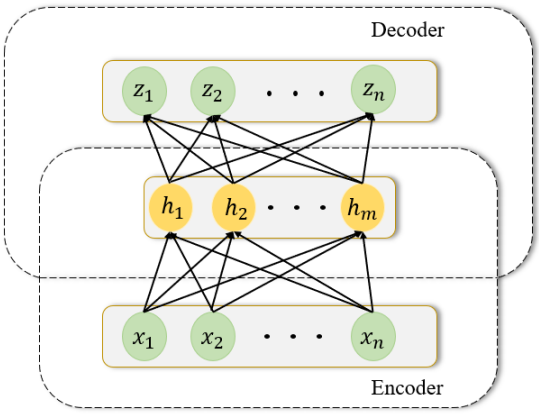
$$sf(t) = 1/\left(1 + e^{-t}\right) \tag{2.6}$$

sendo $\mathbf{W}^{(1)}$ a matriz de pesos associados aos neurônios de entrada e $\mathbf{b}^{(1)}$ o vetor de bias da camada de entrada. Após etapa do *encoder* é necessária a reconstrução dos dados para se encontrar o vetor de saída $\mathbf{z}$:

$$\mathbf{z} = sf\left(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)}\right) \tag{2.7}$$

em que $\mathbf{W}^{(2)}$ é a matriz de pesos associados aos neurônios de saída e $\mathbf{b}^{(2)}$ o vetor de *bias*.

Figura 2.12 – Estrutura de um *autoencoder*.



Fonte: próprio autor.

As funções de otimização dos *autoencoders* são apresentadas em Lu et al. (2016), Abdellatif et al. (2018). Elas são aplicadas para otimizar os parâmetros $\theta = \{\mathbf{W}^1, \mathbf{b}^1, \mathbf{W}^2, \mathbf{b}^2\}$ na construção do *autoencoder*. A função de custo a ser minimizada $E(\theta)$ durante a otimização dos parâmetros da rede é formada por três parcelas:

$$E(\theta) = J_{MSE}(\theta) + J_{Sparse}(\theta) + J_{weight}(\theta). \tag{2.8}$$

A primeira parcela é definida pelo erro médio quadrático MSE (do inglês, *Mean Square Error*) de um *autoencoder*, apresentada por Wen, Gao e Li (2019):

$$J_{MSE}(\theta) = \frac{1}{n}\sum_{i=1}^{n} L_{MSE}(x_i, z_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\|(x_i - z_i)\|^2\right) \tag{2.9}$$

sendo n o número de amostras disponíveis.

Dada uma amostra de entrada $\mathbf{x}$, na qual ρ_j $(j = 1, ..., m)$ é a ativação média da unidade oculta j, que compõe a segunda parcela da função de otimização, que pode ser definida por Wen, Gao e Li (2019), Lu et al. (2016):

$$J_{Sparse}(\theta) = \beta \sum_{j=1}^{m_2} KL(\rho, \hat{\rho}_j), \tag{2.10}$$

sendo,

$$KL(\rho, \hat{\rho}_j) = \rho \log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}, \tag{2.11}$$

e

$$\hat{\rho}_j = \frac{1}{n}\sum_{i=1}^{n}\left[h_j(x_i)\right], \tag{2.12}$$

em que β é o parâmetro de ajuste de peso, que determina a proporção de dispersividade empregada no processo de representação esparsa, m_2 é o número de neurônios na segunda camada, $\hat{\rho}_j$ é o valor médio de ativação para a j-ésima unidade de camada escondida, ρ é o parâmetro de dispersividade e n refere-se ao número de entradas. Observa-se que um termo a mais foi adicionado na divergência de Kullback–Leibler (KL) que penaliza $\hat{\rho}_j$ ao se desviar significativamente de ρ conforme formulado em Lu et al. (2016).

Por fim, para evitar o *overfitting* também há um termo de decaimento que é somado aos demais termos para se encontrar a função de erro de um *autoencoder*:

$$J_{weight}(\theta) = \frac{\lambda}{2}\sum_{l=1}^{2}\sum_{i=1}^{S_l}\sum_{j=1}^{S_{l+1}}\left(w_{ij}^{(l)}\right)^2 \tag{2.13}$$

em que λ é um termo de regularização para ajudar a evitar o *overfitting*, diminuindo a magnitude dos pesos e S_l denota o número de neurônios totais na camada l.

2.3.2 Reconhecimento de Padrões

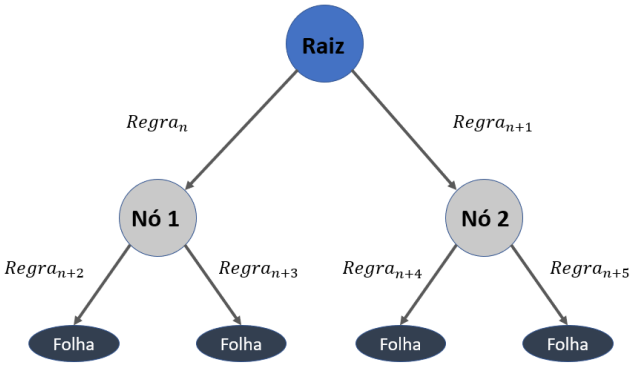
Neste trabalho são empregadas técnicas comumente utilizadas em reconhecimento de padrões, que é o ato de coletar dados brutos e tomar uma ação baseada na 'categoria' do padrão (DUDA et al., 2001). Suas etapas de desenvolvimento podem ser da seguinte ordem: aquisição dos dados, pré-processamento, extração de características e classificação. Os métodos abordados na dissertação são: Árvore de Decisão, Análise de Discriminante Linear, Máquina de Vetores de Suporte, K vizinhos mais próximos e Floresta de Isolamento.

2.3.2.1 Árvore de Decisão

Árvores de Decisão estão condicionadas ao subcampo de aprendizagem de máquina. Isso se deve à sua habilidade de aprender, por meio de exemplos, decompondo um problema complexo em subproblemas simples, aplicando essa estratégia de forma recursiva (BREIMAN et al., 1984).

Brandmaier et al. (2013) diz que as árvores de decisão são estruturas hierárquicas de regras de decisão que descrevem diferenças em um resultado com relação às variáveis observadas.

Figura 2.13 – Estrutura de uma árvore de decisão, n =1.



Fonte: próprio autor.

De acordo com as regras pré estabelecidas, são tomadas decisões de qual caminho seguir, sendo uma estratégia de dividir para conquistar. Conforme Figura 2.13, o nó Raiz executa a primeira divisão, os Nós 1 e 2 são sub-nós, onde ocorrem as decisões e a Folha é local que se estima o resultado, não ocorrendo mais sub-divisões. Em cada nó ou sub-nó há regras a serem seguidas, onde se determina qual direção a árvore irá assumir.

2.3.2.2 Análise de Discriminante Linear

A análise discriminante é uma técnica de estatística multivariada, utilizada para discriminar e classificar objetos. Tem por objetivo estudar o modo de separação de objetos de uma população em duas ou mais classes (KHATTREE; NAIK, 2000).

Esse método consiste em obter funções matemáticas com capacidade para classificar um indivíduo em uma de várias populações. Essa classificação é realizada com base em determinadas características, buscando minimizar as chances de classificar erroneamente um indivíduo em uma população à qual não pertence (NAVES, 2015).

Uma variável aleatória **x** pode ser incluída em uma de **J** classes, com densidade de probabilidade $f_j(\mathbf{x})$. A regra discriminatória tenta dividir o espaço de dados em **J** regiões que representam as classes. A análise de discriminante tem como objetivo alocar **x** à classe j se **x** está na região de j (IZENMAN, 2008).

2.3.2.3 Máquina de Vetores de Suporte

As máquinas de vetor de suporte (SVM, do inglês: *Support Vector Machine*) são modelos de aprendizado supervisionado que podem ser usados para classificação ou detecção. Elas são classificadores de margem, que encontram um hiperplano para decidir a classe para um novo ponto do conjunto de dados (SCHLAG; SCHMITT; SCHULZ, 2019).

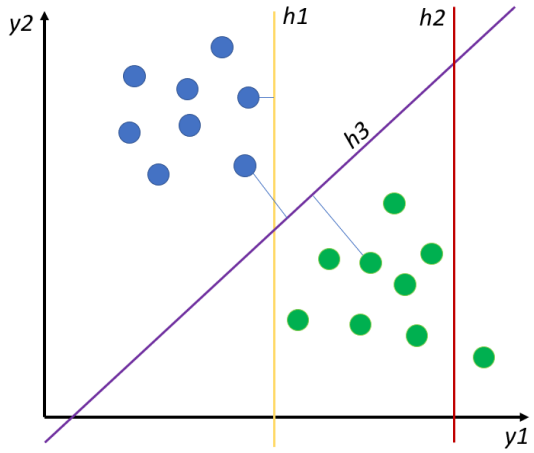
Nguyen (2016) supõe que existam vetores p-dimensionais. Cada um deles pertence a uma das duas classes. Podem-se encontrar muitos hiperplanos dimensionais p–1 que classificam esses vetores, mas há apenas um hiperplano que maximiza a margem entre duas classes. Em outras palavras, o mais próximo entre um lado deste hiperplano e o outro lado é maximizado. Este, é chamado de hiperplano de margem máxima e é considerado como o classificador SVM (NGUYEN, 2016). Como exemplo na Figura 2.14, são gerados três hiperplanos h1, h2 e h3. Com a separação por hiperplanos, somente h3 melhor diferencia as duas classes.

Já o algoritmo *one-class* SVM (OCSVM) é um algoritmo de classificação de apenas uma classe, sendo uma adaptação proposta por Schölkopf et al. (2001). O conceito do *one-class* SVM consiste em encontrar uma hiperesfera em que a maioria das amostras de treinamento sejam incluídas em um volume mínimo (GUERBAI; CHIBANI; HADJADJI, 2014).

O algoritmo OCSVM primeiro mapeia os dados de entrada em um espaço de recursos de alta dimensão por meio de uma função *kernel* e, em seguida, encontra iterativamente o hiperplano de margem máxima, que separa melhor os dados de treinamento da origem. Assim, o

hiperplano (ou limite de decisão linear) corresponde à função de classificação (MAGLARAS; JIANG, 2014).

Figura 2.14 – Separação por hiperplanos.



Fonte: próprio autor.

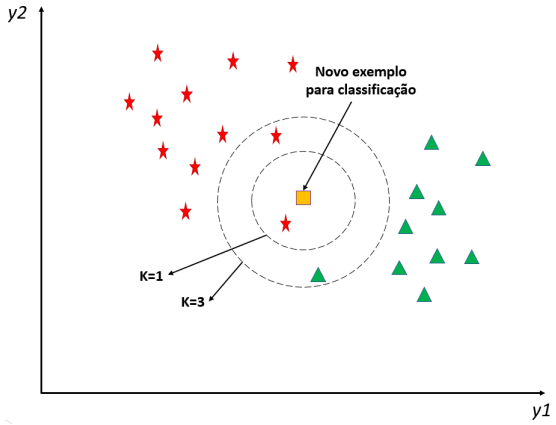
2.3.2.4 K Vizinhos mais Próximos

O método k vizinhos mais próximos (k-NN, do inglês: *k-nearest neighbor*) procura classificar uma amostra desconhecida com base na classificação conhecida de seus vizinhos (MUCHERINO; PAPAJORGJI; PARDALOS, 2009). Essa é uma técnica útil, na qual se atribui pesos às contribuições dos vizinhos para que os vizinhos mais próximos contribuam mais com a média do que os vizinhos mais distantes (WYLD et al., 2011).

Os métodos de vizinhos mais próximos atribuem um valor previsto a uma nova observação com base na pluralidade ou média de seus k "vizinhos mais próximos" no conjunto de treinamento. Dada uma quantidade infinita de dados, qualquer observação terá muitos "vizinhos" arbitrariamente próximos em relação a todas as características medidas e a variabilidade de seus resultados fornecerá uma previsão tão precisa quanto teoricamente possível (RICHMAN, 2011).

A Figura 2.15 mostra um novo exemplo para classificação, onde k representa a variável da quantidade de vizinhos mais próximos. Após a aferição das distâncias, a classe que for predominante dentre as amostras de treinamento será a rotulação do novo exemplo de treinamento.

Figura 2.15 – Ilustração de uma classificação por método k-NN.

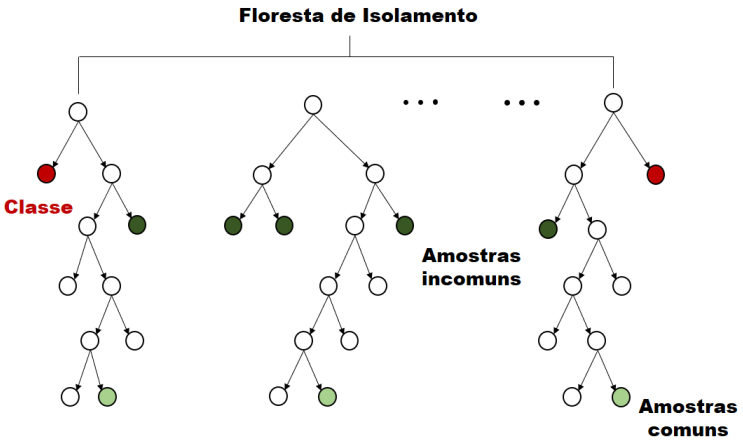


Fonte: próprio autor.

2.3.2.5 Floresta de Isolamento

A Floresta de Isolamento é um algoritmo de aprendizagem não supervisionado para detecção de anomalias que funciona com base no princípio de isolar anomalias (LIU; TING; ZHOU, 2008), em vez das técnicas mais comuns de criação de perfil de pontos normais (CHANDOLA; BANERJEE; KUMAR, 2009). É um algoritmo baseado em árvore de decisão tipo *ensemble*, que constitui uma floresta. Uma técnica é chamada de *ensemble* quando um conjunto de classificadores é treinado individualmente mas as decisões são tomadas de forma combinada (JUNIOR et al., 2020). Métodos *ensemble* tendem a apresentar um menor *overtting* (AGGARWAL; SATHE, 2017).

Figura 2.16 – Floresta de Isolamento.



Fonte: Adaptado de Junior et al. (2020).

O princípio da Floresta de Isolamento é obter uma estrutura de árvores aleatórias para isolar um tipo de classe do seu conjunto de dados como pode ser observado na Figura 2.16. As anomalias tem maior suscetibilidade ao isolamento e ficam mais perto das raízes das árvores, enquanto os pontos normais são mais difíceis de isolar e geralmente estão no extremo mais profundo da árvore (JUNIOR et al., 2020).

O conjunto de treinamento deve ser composto por instâncias de apenas um tipo de classe, pois classes diferentes no treinamento tendem a abaixar a qualidade da função de decisão. Este método é, portanto, adequado para classificação *one-class* (KRAWCZYK et al., 2017).

2.4 Detecção de Falhas em Sistemas Dinâmicos

Nas indústrias, sempre se almeja buscar a maior disponibilidade dos equipamentos, sendo que maior número de falhas pode acarretar maior tempo de parada dos processos. De acordo com a criticidade ou importância do processo, a manutenção da confiabilidade dos equipamentos e sistemas por meio da detecção de falhas, torna esta ferramenta eficaz na continuidade da produção industrial.

A confiabilidade de um item corresponde à sua probabilidade de desempenhar o seu propósito especificado por um determinado período de tempo e sob condições ambientais pré-determinadas (LEEMIS, 1995). Já a disponibilidade é a capacidade de um item, mediante a manutenção apropriada, de desempenhar sua função requerida em um determinado instante de tempo ou em um período de tempo pré-determinado (FOGLIATO; RIBEIRO, 2009).

A detecção de falhas busca expor possíveis desvios apresentados em um processo a partir de suas variáveis monitoradas. Com o advento de sistemas de medição, os valores das variáveis de processos puderam ser obtidos com maior precisão permitindo um monitoramento mais facilitado (DAI; LIU; LONG, 2008). Para a detecção de anomalias, são de suma importância os procedimentos de identificação, diagnóstico e recuperação do processo (ALMEIDA, 2010).

A identificação da falha consiste em encontrar as variáveis mais importantes para a execução do diagnóstico (BAGAJEWICZ, 2009). O propósito da detecção de falhas, tendo o operador como foco central no monitoramento das variáveis pertinentes do processo, é que, em caso de disfunções dos elementos presentes nos processos, ele possa eliminar os efeitos da falha de forma eficiente (ALMEIDA, 2010).

Na fase de projeto de um sistema de detecção de falhas, deve-se conhecer o sistema que deverá ser implantado para que haja melhor alocação dos sensores na indústria. Sensores são essenciais nos processos industriais em várias tarefas, dentre elas, monitoramento de variáveis, controle automático, detecção de falhas, etc.

Para otimizar os custos, é preciso tornar as plantas industriais mais seguras, melhorar o desempenho operacional e aumentar a produção. A partir desses conceitos, detectores automáticos de falhas podem contribuir para melhoria contínua dos sistemas e processos.

Desde o começo da produção em manufatura, falhas causam transtornos ao processo industrial, gerando perdas de produção ou na segurança das instalações e das pessoas envolvidas no processo produtivo. Os primeiros trabalhos sobre detecção de falhas foram baseados no uso de observadores advindos da modelagem da dinâmica dos processos industriais, nos quais as variáveis medidas são monitoradas a fim de se detectar as falhas (ABREU, 2012).

Isermann e Ballé (1997) classificam as falhas em sistemas dinâmicos de três formas:

1. Falhas abruptas: esse tipo de falha provoca rapidamente desvios nas condições de operação normal dos processos;

2. Falhas incipientes: ela acontece de maneira lenta, devido à ocorrência de desvios graduais nas condições de operação normal dos processos;

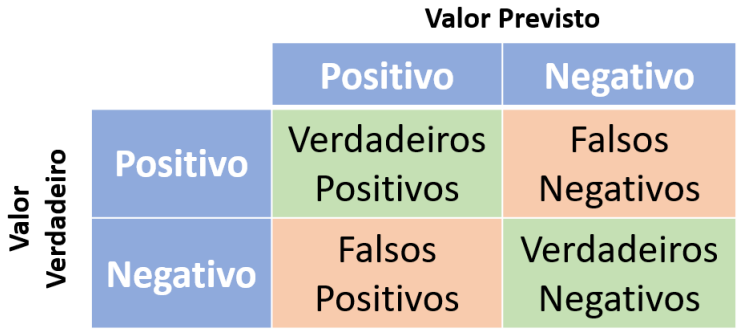
3. Falhas esporádicas ou intermitentes: são falhas que cessam e recomeçam por intervalos, independente do momento.

2.5 Matriz de Confusão

Uma matriz de confusão resume o desempenho de assertividade de um classificador em relação a dados de teste. É uma matriz bidimensional, indexada em uma dimensão pela classe verdadeira de um objeto e na outra pela classe que o classificador atribui (TING, 2017).

Um modelo de classificação é um mapeamento de instâncias com o objetivo de prever classes. Alguns modelos de classificação produzem uma saída contínua (por exemplo, uma estimativa da probabilidade de associação de classe de uma instância) à qual diferentes limiares podem ser aplicados para predizer a participação na classe. Outros modelos produzem um rótulo de classe discreto indicando apenas a classe prevista da instância (FAWCETT, 2006). Um exemplo de matriz de confusão para o caso de duas classes $\{positivo, negativo\}$ é apresentado na Figura 2.17.

Figura 2.17 – Matriz de confusão para duas classes.



Fonte: próprio autor.

A partir da matriz de confusão, algumas métricas podem ser utilizadas para definir o desempenho do classificador:

$$acurracy = \frac{Verdadeiros\,Positivos + Verdadeiros\,Negativos}{Positivos + Negativos} \tag{2.14}$$

$$precision = \frac{Verdadeiros\,Positivos}{Verdadeiros\,Positivos + Falsos\,Positivos} \tag{2.15}$$

$$recall = \frac{Verdadeiros\,Positivos}{Verdadeiros\,Positivos + Falsos\,Negativos} \tag{2.16}$$

$$F1score = \frac{2}{(1/precision) + (1/recall)} \tag{2.17}$$

- *Acurracy*: a acurácia é a exatidão entre os dados classificados de forma correta pelo classificador e o valor verdadeiro dos dados a partir de uma referência pré-estabelecida;
- *Precision*: é a taxa de acerto de uma determinada classe, dividida pelo número de dados classificados para esta classe;
- *Recall*: é a capacidade do modelo em predizer corretamente a condição para casos que realmente a tem;
- *F1 score*: É a média harmônica entre o *recall* e a *precision*. O valor elevado de *F1 score* indica que a acurácia obtida é relevante, não ocorrendo grandes distorções nos dados classificados.

2.6 Validação Cruzada

A validação tem como objetivo verificar se o modelo atende um padrão de desempenho para o qual foi desenvolvido. A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados (KOHAVI, 1995). A generalização do modelo obtido é importante para o trabalho em dados desconhecidos. Dentre as técnicas de validação, as mais conhecidas são: o método *holdout*, o *k-fold* e o *leave-one-out*. No presente trabalho será apresentada a técnica *k-fold*.

O método *k-fold* é uma técnica de validação cruzada, cujas classes são particionadas em tamanhos iguais e exclusivas entre si, sendo k o número de divisões. Após a divisão das classes, cada subconjunto é utilizado para validação e os demais subconjuntos $(k-1)$ são utilizados para estimação dos parâmetros de assertividade, sendo este processo repetido k vezes (RODRIGUEZ; PEREZ; LOZANO, 2010) conforme apresentado na Figura 2.18.

Figura 2.18 – Validação cruzada técnica *k-fold*.

	Fold 1	*Fold 2*	*Fold 3*	*Fold 4*	...	*Fold k*
Iteração 1	Teste	Treino	Treino	Treino	Treino	Treino
Iteração 2	Treino	Teste	Treino	Treino	Treino	Treino
Iteração 3	Treino	Treino	Teste	Treino	Treino	Treino
Iteração 4	Treino	Treino	Treino	Teste	Treino	Treino
⋮	Treino	Treino	Treino	Treino	Teste	Treino
Iteração k	Treino	Treino	Treino	Treino	Treino	Teste

Fonte: próprio autor.

3 MATERIAL E MÉTODOS

Neste capítulo é apresentada a metodologia utilizada no trabalho. Dois conjuntos de dados são empregados para os experimentos. O primeiro é referente a um poço não surgente com elevação artificial por *gas lift* (NASCIMENTO et al., 2020). Já o segundo é de domínio público e foi desenvolvido e disponibilizado por Vargas et al. (2019). São utilizados apenas dados de poços produtores de petróleo reais, não surgentes e surgentes. Desde o tratamento dos dados até a obtenção dos resultados, fez-se uso da ferramenta computacional Matlab®.

3.1 Bases de Dados

3.1.1 Processo não Surgente

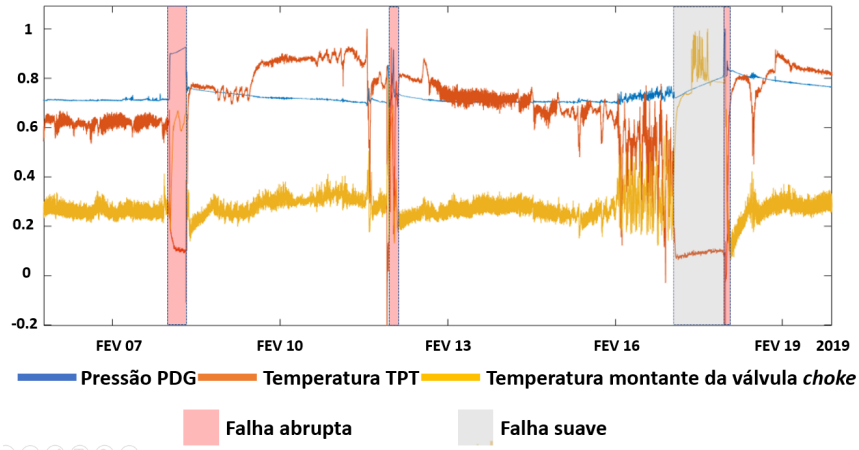
Um dos objetos de estudo deste trabalho são variáveis de poços não surgentes. A aquisição dos dados do poço não surgente operado por *gas lift* foi realizada em uma plataforma de petróleo *offshore*, extraídos do *plant information system* PI *System* da OSIsoft, amplamente utilizado na indústria petrolífera. O PI *System* consiste em um sistema que armazena informações da planta de processo. Ele retorna uma série temporal periódica gerada com interpolação linear entre valores arquivados sem periodicidade definida (VARGAS, 2019). Por meio desse sistema, os dados foram coletados em aproximadamente 90 dias corridos, perfazendo uma janela de observações (que são o conjunto das variáveis em um determinado instante de tempo) de um poço de produção de petróleo. A extração dos dados ocorreu entre os dias 06/12/2018 e 06/03/2019, totalizando 129.592 observações para cada variável, com intervalo de amostragem de um minuto. Campos et al. (2015) explica que, este tempo de amostragem é adequado para o uso pretendido da medição da pressão do poço, ou para a otimização da produção de curto prazo (CAMPOS et al., 2013).

As variáveis utilizadas na concepção dos classificadores estão na Tabela 2.1, totalizando 16 variáveis. Os sensores e atuadores na planta de processo são divididos em variáveis de topo e variáveis de fundo no supervisório. Para as variáveis de topo, os sensores estão localizados na plataforma de produção de petróleo; já para as variáveis de fundo, os instrumentos são instalados no leito do mar e na coluna de produção, podendo ser mais suscetíveis a ruídos e falhas.

Espera-se que o diagnóstico de falha possa ser identificado a partir de certas variáveis inerentes ao processo. Essas variáveis podem indicar a ocorrência de falhas de acordo com as

características intrínsecas dos poços, assim também as condições físicas e químicas podem variar de poço a poço. Para o poço estudado neste trabalho, foi observada entre os dias 16/02/2019 e 19/02/2019 uma falha suave não definida, ou seja, não se sabe a sua origem. Percebeu-se durante a observação que ocorreu a diminuição do fluxo de fluidos, reduzindo lentamente a produção até o cessamento total conforme pode ser observado na Figura 3.1. Já uma falha abrupta é constatada de uma forma mais repentina, alteração os valores das variáveis bruscamente.

Figura 3.1 – Falhas ocorridas entre os dias 07/02/2019 e 19/02/2019.

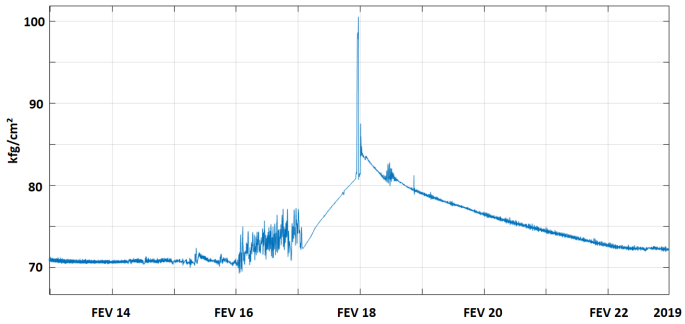


Fonte: próprio autor.

As características de pressão PDG, temperatura na ANM e temperatura a montante da válvula *choke*, de modo a exemplificar uma Falha Suave com estes três sensores, são apresentadas as Figuras 3.2, 3.3 e 3.4. Estas figuras indicam uma anormalidade no comportamento do poço descrito, conforme abaixo:

- Pressão PDG (PT1): tende a ser elevada a partir da ocorrência de uma falha devido a uma obstrução ao longo do caminho do escoamento dos fluidos, elevando a pressão na coluna de produção;

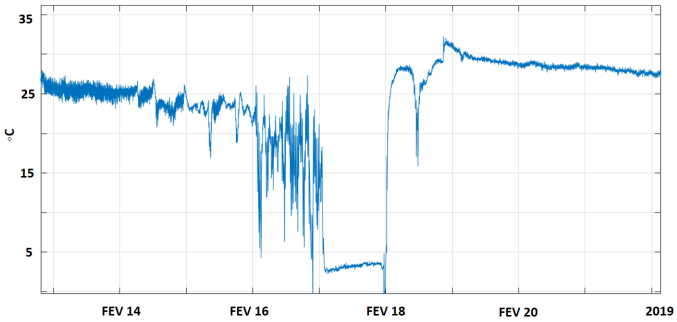
Figura 3.2 – Pressão PDG (PT1).



Fonte: próprio autor.

- Temperatura na árvore de natal molhada (TT2): tende a se equilibrar com a temperatura do leito do mar, no caso de falhas de escoamento de fluidos. Em águas profundas, estas temperaturas variam normalmente entre 2°C e 6°C;

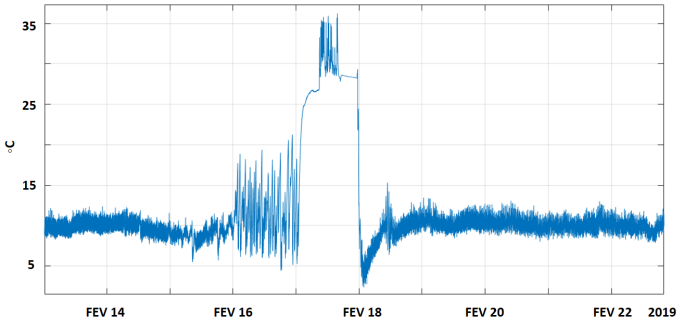
Figura 3.3 – Temperatura na ANM (TT2).



Fonte: próprio autor.

- Temperatura a montante da válvula *choke* (TT3): tende a se igualar à temperatura na superfície do mar.

Figura 3.4 – Temperatura a montante da válvula *choke* (TT3).



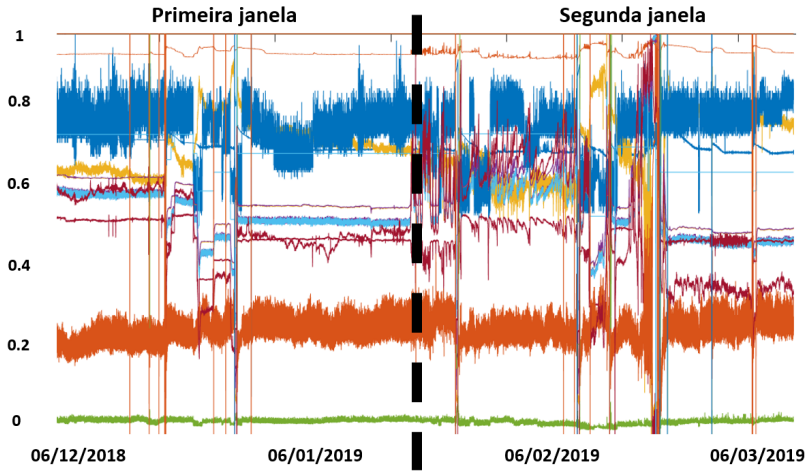
Fonte: próprio autor.

A janela de 129.592 observações foi dividida em duas partes de tamanhos iguais, cada uma com 64.796 observações, indicadas na Figura 3.5 com as 16 variáveis do processo exibidas. Os valores dos sensores nas figuras são normalizados, dentro do intervalo entre zero e um. Para facilitar a visualização dos gráficos e as classes foram definidas como Falha e Não-Falha.

A rotulagem das falhas foi proposta pelos autores, de acordo com informações repassadas por operadores de campo e operadores de sala de controle (supervisório) da produção de poços de petróleo *offshore* com vasta experiência ao longo de suas carreiras na área de óleo e gás. As falhas foram constatadas no supervisório de monitoramento da produção. Na classe Falha há dois padrões de ocorrência: Falha Abrupta e Falha Suave. A quantidade de observações

para a classe Não-Falha é de 126.577 e 3.015 para a classe Falha, sendo que Falha Abrupta e Falha Suave contém 2.143 e 872 observações, respectivamente. Na Tabela 3.1 é apresentada a divisão das classes nas duas janelas de observações.

Figura 3.5 – Conjunto de dados divididos em dois períodos.



Fonte: próprio autor.

Tabela 3.1 – Quantidades de observações rotuladas como Falha e Não-Falha.

Dados por Janela	Falha	Não-Falha
Primeira	498	64.298
Segunda	2.517	62.279

Fonte: próprio autor.

A divisão entre as classes é exibida na Figura 3.1, onde estão plotados os valores dos sensores PT1, TT2 e TT3, no intervalo entre os dias 07/02/2019 e 19/02/2019. Vale salientar que não existe a ocorrência de Falha Suave na primeira janela de dados. Foram utilizados os rótulos 0 ou 1 para treinamento dos classificadores, sendo o alvo a ser alcançado na saída do classificador seja:

- se saída = 0 : Falha;
- se saída = 1 : Não-Falha.

Como em Aguirre et al. (2017), os dados históricos para os poços não surgentes foram obtidos por um sistema SCADA (do inglês, *Supervisory Control and Data Acquisition*) que significa em português (Controle de Supervisão e Aquisição de Dados).

3.1.2 Processo Surgente

Vargas et al. (2019) disponibilizam dados de um conjunto de poços de extração de petróleo *offshore* surgentes chamado 3W *dataset*. Esses dados são advindos do Projeto MAE (Monitoramento de Alarmes Especialistas), da empresa Petróleo Brasileiro SA, concebido na Unidade de Negócios da Petrobras localizada no estado do Espírito Santo (UN-ES). Os dados estão no formato de séries temporais, que consistem em uma sequência de observações ordenadas pelo tempo e que apresentam intervalos de tempo iguais entre cada par de observações.

Os dados são variantes no tempo, nos quais as características físico-químicas dos reservatórios tendem a se alterar ao longo dos dias, meses e anos. Comumente supõe-se que existe correlação entre os dados passados e futuros (MORETTIN; TOLOI, 2006). O conjunto de dados é dividido em oito tipos de eventos que podem causar perdas de produção em conjunto com a operação normal dos poços. A divisão dos dados ocorreu em três tipos de instâncias, que representam ocorrências ou casos, sendo eles:

- Real: dados extraídos extraídos de PI *System* no qual as variáveis ficam gravadas. Na aquisição amostrada foram considerados períodos normais, transitórios e em regime (anomalias). Todos os dados foram analisados e rotulados por especialistas;
- Simulado: as simulações dos cenários de eventos indesejáveis foram geradas por especialistas na área de petróleo, utilizando um *software* de grande aplicabilidade em simulações na indústria do petróleo;
- Desenhado: neste tipo de conjunto de dados, especialistas da área desenharam as variáveis de um evento indesejável em um modelo gráfico, cujos modelos foram impressos em uma folha A4 e a partir desses desenhos digitalizados foram gerados os dados.

A verificação e classificação de falhas em poços de petróleo *offshore* são tarefas complexas, devendo ser realizadas por pessoas qualificadas, ou seja, especialistas em Elevação e Escoamento de petróleo. Esses tipos de dados já pré-tratados auxiliam pessoas não especialistas em trabalhos como de classificação e detecção de anomalias em poços de petróleo. As instâncias utilizadas, que compõem a base de dados do 3W *dataset* estão na Tabela 3.2. Não foram aplicadas as variáveis que tem relação ao sistema de *gas lift*, pois a intenção é o desenvolvimento de detectores de anomalias em poços surgentes. As variáveis integrantes deste processo que integram do banco de dados são:

- Pressão PDG;
- Pressão TPT;
- Temperatura TPT;
- Pressão a montante da válvula *choke* de produção;
- Temperatura a jusante válvula *choke* de produção.

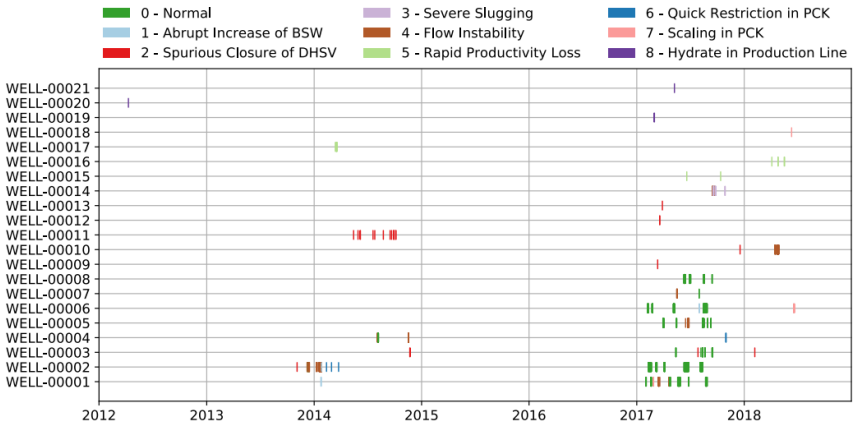
Tabela 3.2 – Quantidades de instâncias que compõem o banco de dados 3W *dataset* (VARGAS et al., 2019).

Tipo de Evento	Real	Simulado	Desenhado	Total
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 na Válvula *Choke*	6	215	-	221
7-Incrustação na Válvula *Choke*	4	-	10	14
8-Hidrato em Linha de Produção	3	81	-	84

Fonte: Adaptado de (VARGAS et al., 2019).

Na Figura 3.6 é apresentado o mapa de dispersão das instâncias reais de acordo com o tipo de evento, em diferentes poços de produção ao longo dos anos, entre 2012 e 2018. As barras coloridas representam o começo e o fim de cada evento. A taxa de amostragem é fixa em um segundo e a quantidade de observações em cada instância varia de acordo com o tipo de evento.

Figura 3.6 – Mapa de dispersão das instâncias reais do conjunto de dados para criação dos modelos.



Fonte: (VARGAS et al., 2019).

3.1.2.1 *Benchmark* Trabalhado

Vargas et al. (2019) propõem dois desafios a partir do banco de dados concebido em artigo divulgado. Esses *benchmarks* são planejados apenas para detecção *online*. Todas as observações de períodos transientes de anomalia e estados estáveis de anomalia devem ser novamente rotulados como positivos e todas as observações de períodos normais como negativas. Nessa operação, observações não rotuladas devem ser mantidas como estão.

Para essa dissertação, apenas um dos desafios propostos (desafio número 2) foi implementado. Vargas (2019) demonstra as três regras que devem ser seguidas para construção dos detectores. Essas regras são transcritas a seguir:

"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 deve ser utilizada 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 *F1* 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 *F1* deve ser considerado a principal métrica de desempenho por estabelecer uma relação de compromisso entre precisão e sensibilidade."

3.2 Construção do Modelos

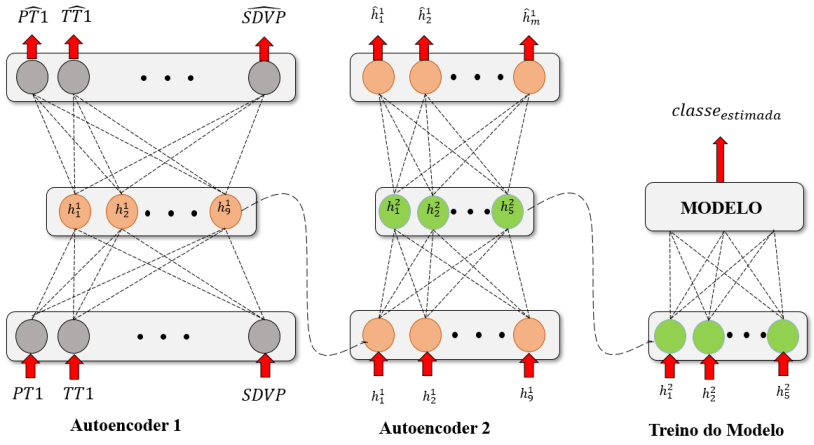
São analisados dois tipos de variáveis: variáveis reais de poços surgentes e variáveis reais de poços não surgentes. Assim, a construção de classificadores é realizada com metodologias diferentes, apresentados nas próximas duas seções.

3.2.1 Construção do Modelo de Poço não Surgente

Na construção dos detectores (ou modelos) de falha no poço não-surgente são empregados dois *autoencoders* empilhados, basseado na abordagem apresenta por Lu et al. (2016) . O primeiro tem a camada escondida com 9 neurônios e o segundo com 5 neurônios. O treinamento do primeiro *autoencoder* é realizado com as 16 variáveis contidas na Tabela 2.1 como entrada e o treino do segundo *autoencoder* utiliza a saída da camada oculta do primeiro *autoencoder* como entrada.

A saída da camada oculta do segundo *autoencoder* é empregada como entrada para o treinamento dos detectores, ou seja, os *autoencoders* são utilizados como pré-processamento dos dados e redução da dimensionalidade de 16 para 5 variáveis de entrada. Foi definida previamente pelos autores uma redução de dimensionalidade em aproximadamente 70% da quantidade inicial de variáveis de entrada com a intenção de reduzir a complexidade computacional do treinamento dos detectores.

Figura 3.7 – Processo de construção de um modelo *stacked autoencoders* para classificação de falhas em poços de produção com elevação por *gas lift*.

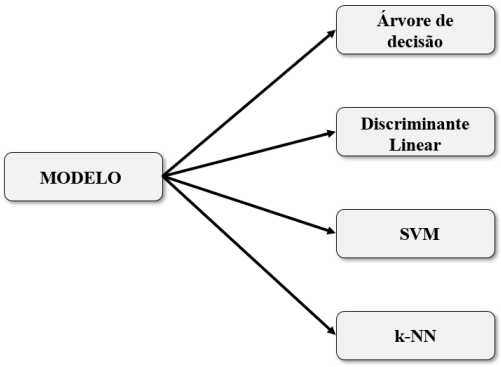


Fonte: próprio autor.

Como observado na Figura 3.7, que apresenta o processo de construção do modelo proposto, onde $v = \{PT1, TT1, ..., SDVP\}$ representa as variáveis de entrada do modelo. Contidas na Tabela 2.1 do processo de extração de petróleo. $\hat{v} = \{\hat{PT1}, \hat{TT1}, ..., \hat{SDVP}\}$ são as variáveis de entrada estimadas na saída dos *autoencoders*, $h^1 = \{h_1^1, h_2^1, ...h_9^1\}$ são os valores de saída da camada oculta do primeiro *autoencoder* e $\hat{h}_1 = \{\hat{h}_1^1, \hat{h}_2^1, ...\hat{h}_9^1\}$ a saída estimada de h^1 no segundo *autoencoder*. Os valores da camada escondida do segundo *autoencoder* são $h^2 = \{h_1^2, h_2^2, ...h_5^2\}$. Eles são os parâmetros de entrada dos detectores testados conforme Figura 3.8, que têm como

saída a classe estimada. Foi criado um modelo k-NN, sendo feita a redução de dimensionalidade aplicando-se a técnica PCA com a utilização de 5 componentes, a mesma quantidade proposta com *autoencoders*. Também são criados modelos sem a redução de dimensionalidade, utilizando as 16 características de entrada disponíveis para comparação do custo computacional e acurácia da técnica proposta.

Figura 3.8 – Modelos testados na classificação de falhas em poços de produção com elevação por *gas lift*.



Fonte: próprio autor.

Neste trabalho, são desenvolvidos quatro modelos comumente empregados em reconhecimento de padrões: Árvore de Decisão, Análise de Discriminante Linear, Máquina de Vetores de Suporte e K vizinhos mais próximos. Para a análise do desempenho dos classificadores, são utilizadas as métricas *recall*, *precision* e *F1 score*.

Para o desenvolvimento dos modelos propostos, são empregados classificadores disponíveis na ferramenta *Classification Learner*, aplicativo do *software* Matlab®. Para a criação dos modelos, os hiperparâmetros foram ajustados na própria ferramenta, sendo exibidos na Tabela 3.3 de forma que o desempenho de cada classificador .

A técnica de validação cruzada é aplicada para avaliar a capacidade de generalização dos modelos a partir de um conjunto de dados (KOHAVI, 1995). Dentre as técnicas de validação as mais utilizadas são *holdout*, *k-fold* e *leave-one-out*, sendo a última utilizada neste trabalho.

No método *k-fold* os dados são divididos em parcelas (*folds*) de tamanhos iguais e exclusivas entre si, sendo k o número de divisões. Após a divisão dos dados em cada subconjunto, uma partição é utilizada para validação e as demais $(k-1)$ são utilizadas para estimação dos parâmetros, sendo este processo repetido k vezes a fim de se obter o desempenho do modelo.

São empregados dois *folds*. Nas observações utilizadas para treino, cada classe foi particionada em dois conjuntos de tamanhos iguais, utilizados para treino, teste e validação. O segundo conjunto de observações é aplicado apenas para teste dos modelos constituídos. Esse processo é repetido uma vez.

Tabela 3.3 – Hiperparâmetros dos modelos desenvolvidos para poços não surgentes.

Modelos	Hiperparâmetros
Árvore de Decisão	Número máximo de divisões: 25 Critério de impureza: Índice Gini
Discriminante	Estrutura de covariância: Total
SVM	Função kernel: Gaussiana Escala Kernel: 0,56 Método Multiclasses: *One-vs-One*
k-NN	Número de Vizinhos: 7 Métrica de distância: Euclidiana Pesos das distâncias: Igual

Fonte: próprio autor.

3.2.2 Construção dos Modelos de Poços Surgentes

Para o processo a ser implementado para a detecção de falhas em poços de petróleo *offshore* surgentes, são utilizadas apenas as variáveis que constam em (VARGAS et al., 2019), no total 5 variáveis. De acordo com as regras propostas por Vargas et al. (2019), foi definida uma sequência de passos na construção dos classificadores (KADHIM, 2019):

- Pré-processamento dos dados;
- Extração de características;
- Aplicação de técnicas de detecção;
- Análise e avaliação dos resultados obtidos.

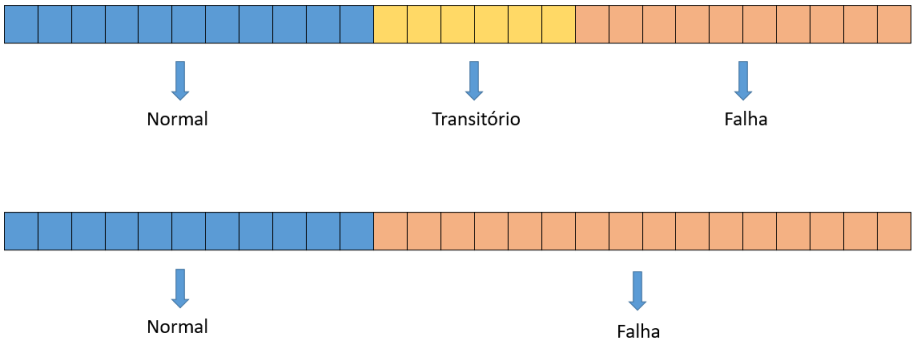
3.2.2.1 Pré-processamento dos Dados

Observações com algum dado NaN (do inglês, *Not a Number*) foram interpoladas. As observações com transiente de anomalia foram re-rotuladas como anomalia, como apresentado na Figura 3.9. Os dados foram normalizados, com média zero e variância unitária, com o objetivo de evitar que a diferença desproporcional entre as escalas interferisse nos resultados.

Atrasos foram inseridos em cada observação, prefazendo um total 100 atrasos, ou seja, $(k-1),(k-2)...(k-100)$, conforme Figura 3.10.

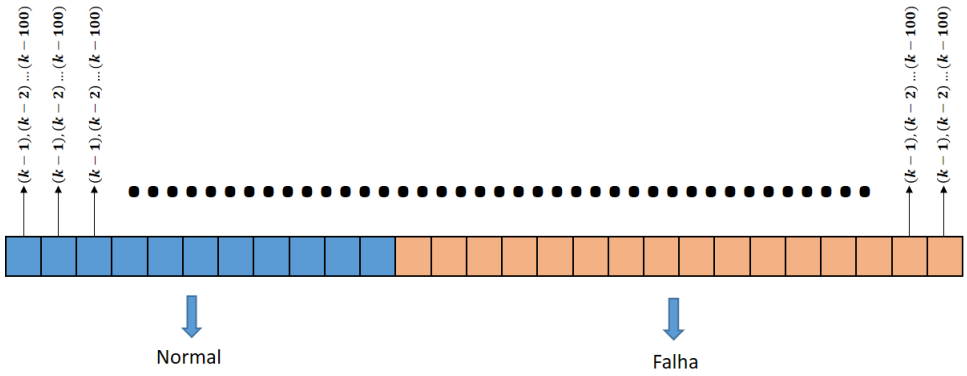
Com este procedimento para as instâncias ficarem homogêneas, são descartadas as primeiras cem observações para treinamento. Após a inclusão dos atrasos, a quantidade de variáveis passou de 5 para 500, ou seja, multiplicadas por 100. Este procedimento aumentou significativamente o volume de dados para o treinamento e teste.

Figura 3.9 – Dados de transitório re-rotulados em falha.



Fonte: próprio autor.

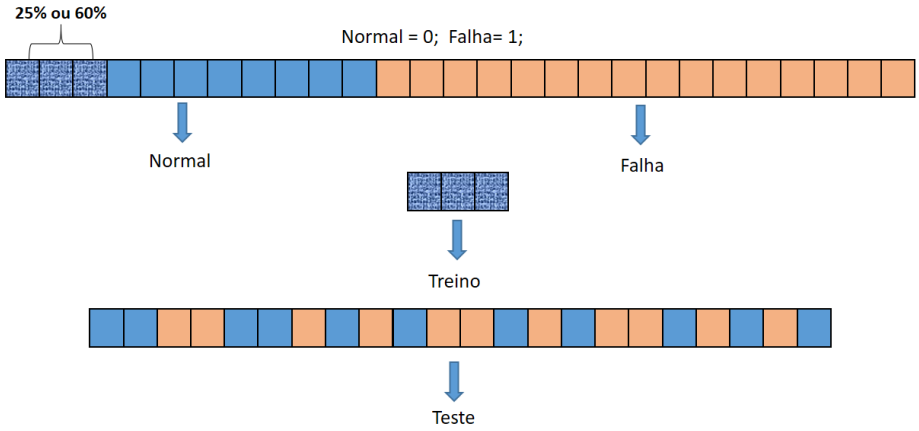
Figura 3.10 – Atrasos nas observações.



Fonte: próprio autor.

Em períodos normais, as primeiras observações foram utilizadas para treinamento (60%) e as últimas para validação (40%). Os 40% restantes de amostras normais são misturados de forma aleatória com as amostras de anomalia. Esta partição dos dados normais foi realizada nos trabalhos de Vargas et al. (2019) e Junior et al. (2020). Também foi empregado a utilização de (25%) dos dados normais para treinamento, conforme Figura 3.11, que podem auxiliar na análise deste parâmetro.

Figura 3.11 – Observações de treino e teste.

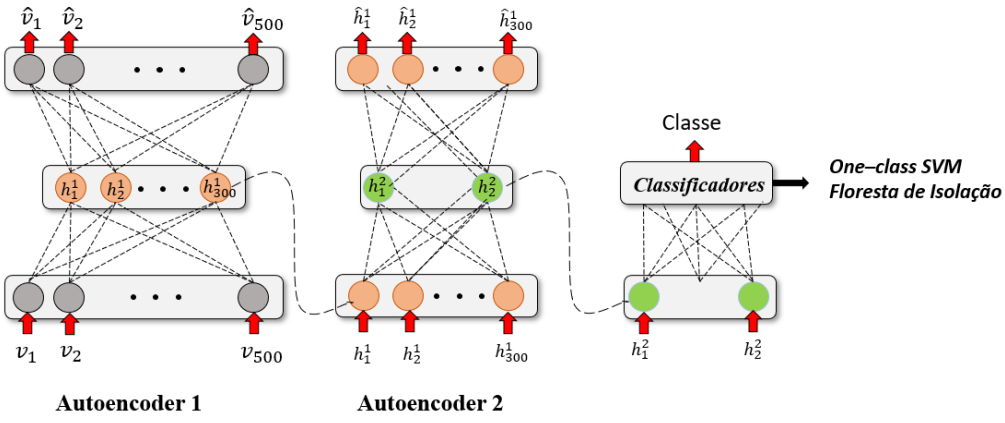


Fonte: próprio autor.

3.2.2.2 Redução de Dimensionalidade

Após o pré-processamento dos dados, o processo de redução de dimensionalidade é realizado por meio de *autoencoders* em cascata. A Figura 3.12 mostra o processo de construção do modelo, utilizando 500 variáveis de entrada, onde $v = \{v_1, v_2, ...v_{500}\}$ são as variáveis de entrada do modelo. $\hat{v} = \{\hat{v_1}, \hat{v_2}, ...\hat{v_{500}}\}$ são as variáveis de entrada estimadas na saída dos *autoencoders*, $h^1 = \{h_1^1, h_2^1, ...h_{300}^1\}$ são os valores de saída da camada oculta do primeiro *autoencoder* e $\hat{h}^1 = \{\hat{h}_1^1, \hat{h}_2^1, ...\hat{h}_{300}^1\}$ a saída estimada de h^1 no segundo *autoencoder*. Os valores da camada escondida do segundo *autoencoder* são $h^2 = \{h_1^2, h_2^2\}$ e também servem como parâmetros de entrada dos detectores.

Figura 3.12 – Redução de dimensionalidade e classificação.



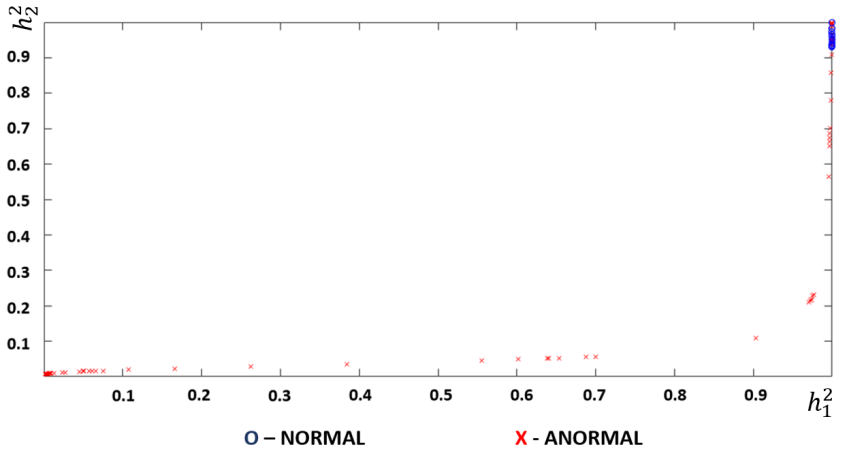
Fonte: próprio autor.

Como exemplo, as Figuras 3.13 e 3.14 apresentam as duas classes (Normal x Anormal) após a redução da dimensionalidade por *autoencoders* do poço WELL00006201806180602 45,

a segunda mostra a imagem ampliada da zona onde há a transição entre normalidade e anormalidade. Também são apresentados os gráficos do poço WELL00021201705090135 17 nas Figuras 3.15 e 3.16. h_1^2 e h_2^2 são as saídas da camada oculta do segundo *autoencoder*. Os gráficos foram gerados com apenas parte do banco de dados do poço, espaçado em 50 em 50 unidades.

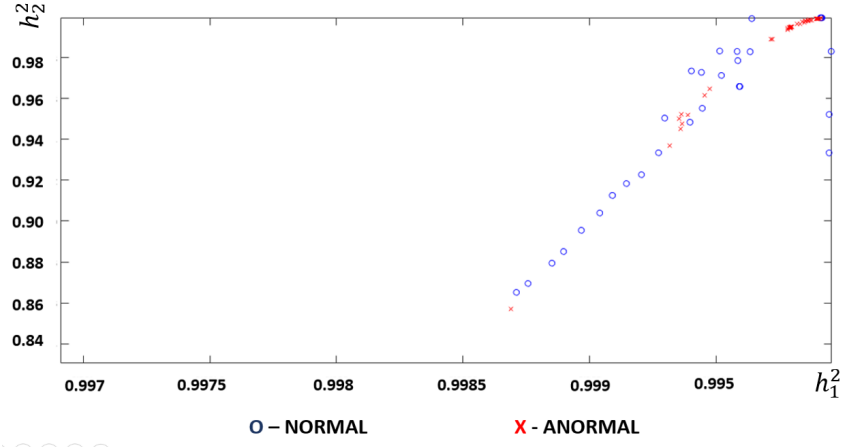
São criados dois modelos Floresta de Isolamento com a redução de dimensionalidade aplicando a técnica PCA. Para comparação o desempenho das duas técnicas de redução de dimensionalidade (*autoencoders* x PCA). São reduzidos de 500 variáveis para 100 e de 500 para 10. Os detectores são implementados por meio do algoritmo de Floresta de Isolamento. Para verificar a eficiência dos modelos acrescidos de atrasos também é treinado um modelo de Floresta de Isolamento sem inclusão de atrasos.

Figura 3.13 – Poço WELL00006201806180602 45 redução de dimensionalidade.



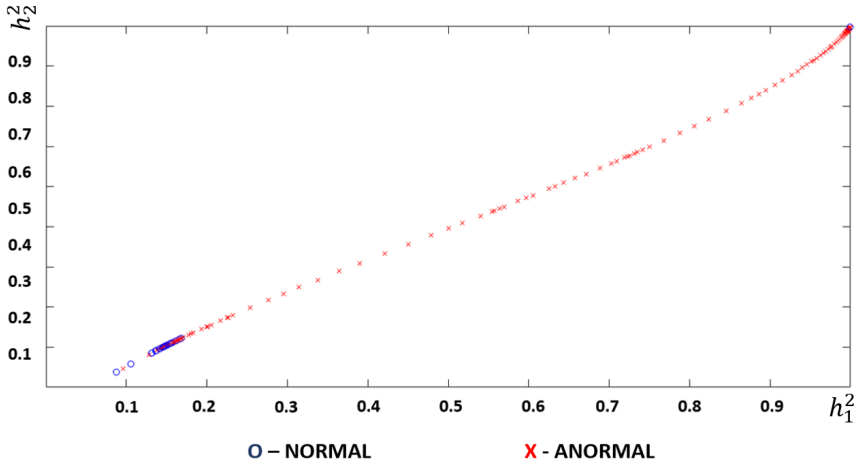
Fonte: próprio autor.

Figura 3.14 – Zoom poço WELL00006201806180602 45 após redução de dimensionalidade.



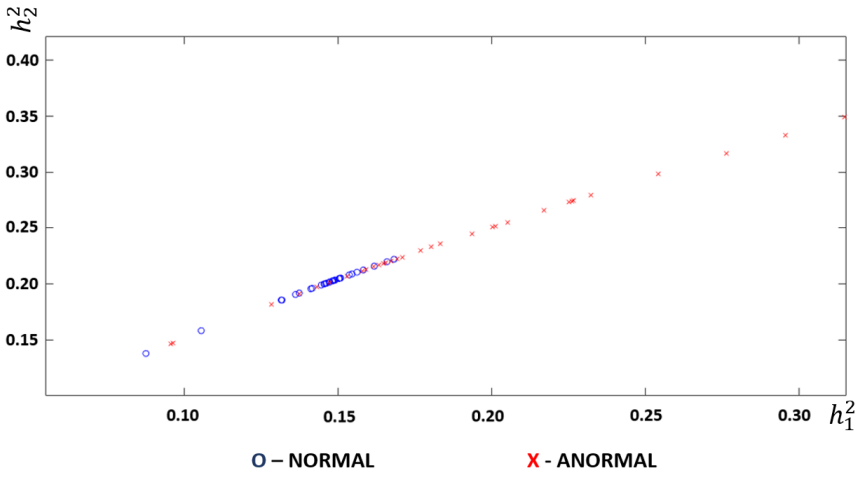
Fonte: próprio autor.

Figura 3.15 – Poço WELL00021201705090135 17 após redução de dimensionalidade.



Fonte: Autor

Figura 3.16 – Zoom poço WELL00021201705090135 17 após redução de dimensionalidade.



Fonte: próprio autor.

3.2.2.3 Aplicação de Técnicas de Detecção

Foram utilizadas duas técnicas para detecção de anomalias, SVM *one-class* e Floresta de Isolamento, com código disponibilizado por Liu, Ting e Zhou (2008). Nos trabalhos de Vargas et al. (2019) e Junior et al. (2020) também são aplicadas essas técnicas e, por esse motivo, são escolhidas para criação dos modelos. Os hiperparâmetros dos detectores desenvolvidos neste trabalho não passaram por processo de otimização, eles foram retirados do trabalho de Junior et al. (2020), onde já foram otimizados. Os hiperparâmetros dos detectores desenvolvidos podem ser consultados na Tabela 3.4.

Após o desenvolvimento dos modelos, os resultados de detecção obtidos são apresentados no próximo capítulo.

Tabela 3.4 – Hiperparâmetros dos modelos desenvolvidos para poços surgentes.

Modelos	Hiperparâmetros
Floresta de Isolamento	n estimators: 150 max samples: 1,0 max featuress: 1,0 bootstrap: False contamination: 0
SVM *one-class*	Função kernel: rbf gama: 0,001 nu: 0,1

Fonte: próprio autor.

4 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados resultados e discussões dos experimentos e análises realizadas. Primeiramente são apresentados os resultados dos dados do poço não-surgente; logo após, são apresentados os resultados dos poços surgentes.

4.1 Poço não Surgente

Os modelos são implementados com classes tendo quantidades diferentes de observações como apresentado na Seção 3.1.1. Apesar desta discrepância, a maioria dos modelos classificaram de forma satisfatória o conjunto de observações.

As Tabelas 4.1 e 4.2 apresentam o desempenho dos modelos em dados de teste das métricas *recall*, *precision* e *F1 score*. Nessas tabelas são apresentados os resultados dos modelos SVM e k-NN com todas as 16 entradas, ou seja, sem implementação dos *autoencoders*, pois esses classificadores obtiveram um desempenho superior na métrica *F1 score*.

Ao utilizar o primeiro grupo de dados para treino, obteve-se um resultado inferior de *F1 score* em dados de teste com relação ao resultado obtido quando utilizada a segunda janela de observações para treino. Uma análise das possíveis causas para este fato, é devido ao número reduzido de observações da classe Falha no primeiro conjunto de dados. Também por este conjunto não possuir observações de Falha Suave, que são menos suscetíveis de serem detectadas.

Os modelos k-NN obtiveram, de forma geral, desempenhos superiores às outras técnicas nas métricas calculadas. Esta superioridade é observada especialmente no teste do modelo treinado na primeira janela de dados e testado na segunda.

Tabela 4.1 – Valores das métricas *recall*, *precission* e *F1 score* para os dados de teste da segunda janela de dados.

Modelos	*recall*	*precission*	*F1 score*
Árvore de Decisão	0.7799	0.3758	0.5072
Discriminante	0.9329	0.4529	0.6098
SVM	0.4000	0.0064	0.0125
k-NN	0.9943	0.4819	0.6492
SVM sem redução	0	0	0
k-NN sem redução	0.9956	0.5447	0.7042
k-NN com PCA	0.9862	0.4251	0.5941

Fonte: próprio autor.

Tabela 4.2 – Valores das métricas *recall*, *precission* e *F1 score* para os dados de teste da primeira janela de dados.

Modelos	*recall*	*precission*	*F1 score*
Árvore de Decisão	0.8418	0.8012	0.8210
Discriminante	0.9062	0.7952	0.8471
SVM	0.7878	0.8574	0.8212
k-NN	0.8846	0.8313	0.8571
SVM sem redução	0.8692	0.8421	0.8554
k-NN sem redução	0.9195	0.8996	0.9095
k-NN com PCA	0.8634	0.8176	0.8398

Fonte: próprio autor.

Apesar do modelo k-NN sem aplicação dos *autoencoders* ter alcançado melhores métricas em relação à sua versão com redução de dimensionalidade, observa-se que no quesito *recall* esta diferença no desempenho é ainda menor. Para o problema deste trabalho, essa técnica é mais importante, pois a presença de falso negativo é mais problemática. O custo de um falso negativo em geral é maior. Um evento anormal classificado como normal é mais prejudicial do que um evento normal classificado como anormal.

Portanto, além do fato do uso de *autoencoders* não afetar de sobremaneira os índices de desempenho, a redução de dimensionalidade propicia uma diminuição do tempo de execução dos modelos. Quando utilizada a técnica k-NN, esse fator está na ordem de aproximadamente oito vezes. A taxa de execução é de 12.000 observações por segundo para o modelo com redução e 1.600 observações por segundo sem redução, em um computador com processador Intel® Core i3 ™ e quatro gigas de memória RAM. O tempo de execução do algoritmo de redução de dimensionalidade *autoencoder* é inferior a vinte segundos, sendo portanto irrelevante na soma tempo de total na execução dos modelos.

Já os modelos treinados com a técnica SVM, quando comparados entre si, têm resultados próximos. Utilizando a segunda janela para treinamento e a primeira para teste, o desempenho é nulo no modelo com todas as variáveis de entrada. Apesar do bom desempenho em dados de treino, toda a classe Falha foi classificada como Não-Falha, ou seja, todos falsos negativos. Mesmo com os resultados da primeira janela sendo elevados, o desempenho deste modelo é muito inferior aos demais. Já no outro experimento, os modelos SVM obtiveram desempenho acima de 0.85 na métrica *F1 score*.

Os classificadores baseados em árvore de decisão e discriminante conseguiram resultados acima de 0.50 da métrica *F1 score* em ambos os testes. Como todos os classificadores

propostos, o desempenho dessas duas técnicas é inferior quando os modelos são testados na segunda janela de observações. O modelo discriminante conseguiu atingir um elevado valor de *recall* nessa janela, 0.90. O uso *autoencoder* foi recomendado para diminuir o custo da execução de tarefas computacionais e o modelo k-NN foi o melhor classificador dentre os testados para este tipo de série temporal e falha analisada.

A técnica PCA obteve valores de desempenho inferiores aos apresentados quando utilizados *autoencoders* para este tipo de conjunto de dados, o que representa a viabilidade dos *autoencoders* em problemas que necessitem da redução de dimensionalidade. Para o caso de um poço não surgente a redução de dimensionalidade não foi de uma ordem elevada em relação ao número de características, ficando por volta de um terço da quantidade de variáveis. Já para o próximo tópico, onde são analisadas anomalias em poços surgentes, a redução por *autoencoder* se torna mais significativa.

4.2 Poços Surgentes

O conjunto de dados disponibilizados por (VARGAS et al., 2019) foi utilizado na criação de detectores de anomalia. A partir dos resultados obtidos, são usados os modelos propostos por (VARGAS, 2019) e (JUNIOR et al., 2020) para avaliação e comparação dos seus desempenhos. As regras impostas no *benchmark* estão descritas na Subseção 3.1.2.1.

Tanto no trabalho de Junior et al. (2020) como em Vargas (2019) é utilizada a estratégia específica de amostragem com janela deslizante. Também foram realizadas amostragens das instâncias com janela deslizante com a geração de até 15 amostras com 180 observações cada. Os autores extraíram e utilizaram como características: mediana, média, desvio padrão, variância, máximo e mínimo.

As etapas de implementação dos modelos desta dissertação estão na Subseção 3.2.2. Foram aplicados dois tipos de detectores de uma única classe: Floresta de Isolamento e SVM *one-class*. Foram realizas 31 rodadas de treinamento e validação para cada modelo, totalizando 434 avaliações conforme *benchmark* proposto. Os valores obtidos de *recall*, *precision* e *F1 score* dos modelos de floresta de isolamento e SVM *one-class* podem ser vistos nas tabelas que constam no Apêndice A.

Quando analisados múltiplos detectores pode-se gerar métricas *F1* cujas médias sejam iguais entre si. Assim foi realizado o Teste de Friedman (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Aplicado nos algorítimos experimentados dos modelos com 60% de observações

normais para treinamento. Com o resultado deste teste de p-*value* 2,5x10^{-11}. A hipótese de todos os algoritmos terem médias de *F1* iguais pode ser rejeitada, o mesmo teste foi realizado em Vargas et al. (2019) e Junior et al. (2020).

4.2.1 Análise de Resultados

Vargas et al. (2019) e Junior et al. (2020) apresentam os resultados de seus trabalhos por meio de tabelas que contêm os valores da média *F1* e o desvio padrão dos poços reais testados do conjunto de dados 3W *dataset*.

Foram implementados modelos com (25%) e (60%) dos dados normais para treino do detector. Também foi reduzida a dimensionalidade dos dados para cem, cinquenta, dez e duas dimensões. Os resultados dos modelos obtidos nesta dissertação podem ser vistos na Tabela 4.3, representados por meio da sua média *F1* e respectivo desvio padrão.

Tabela 4.3 – Resultados dados 3W média *F1* e desvio padrão dos modelos desenvolvidos *F1* (Média) ± Desvio Padrão.

Modelo	*one-class* SVM	Floresta de Isolamento
25% e duas dimensões	0,6565 ± 0,2394	0,7865 ± 0,1703
60% e duas dimensões	0,6814 ± 0,2170	0,7997 ± 0,1532
60% e dez dimensões	0,6926 ± 0,2068	0,8067 ± 0,1456
60% e cinquenta dimensões	0,7033 ± 0,2069	0,8160 ± 0,1396
60% e cem dimensões	0,7201 ± 0,1972	0,8277 ± 0,1360
60% e cem dimensões com PCA	-	0,7956 ± 0,1428
60% e dez dimensões com PCA	-	0,7328 ± 0,1507
sem inclusão de atrasos	-	0,6766 ± 0,1395
com 500 atrasos e sem redução	-	0,8345 ± 0,1329
(VARGAS et al., 2019)	0,5320 ± 0,0750	0,7430 ± 0,1790
(JUNIOR et al., 2020)	0,5670 ± 0,1620	0,7270 ± 0,1820

Fonte: próprio autor.

A partir da comparação dos resultados obtidos nota-se que o método com a inserção de atrasos, onde promove uma expansão da quantidade de dados por observação em conjunto com a aplicação de *autoencoders* na redução de dimensionalidade, conseguiu melhores resultados na média *F1 score* em ambos detectores.

Mesmo sem aplicação de técnicas de otimização, os modelos desenvolvidos apresentaram resultados satisfatórios. Os valores de desvio padrão *F1 score* do modelo *one-class* SVM ficaram acima dos demais. Já os resultados do desvio padrão dos modelos implementados com floresta de isolamento ficaram um pouco abaixo dos modelos propostos em outros trabalhos.

Com a aplicação de 25% dos dados normais em cada instância para treinamento houve um pequeno decréscimo no desempenho dos detectores, em comparação com a utilização de 60% dos dados para treino. Observa-se também que o desempenho dos modelos obtidos neste trabalho é superior àqueles apresentados em Vargas et al. (2019) e Junior et al. (2020). O uso de uma janela de 500 atrasos é de suma importância para o desempenho dos classificadores propostos. Além disso, o modelo sem a inclusão de atrasos teve um desempenho inferior quando comparado aos com atrasos e redução de dimensionalidade. Um diferencial do trabalho proposto aos pares que são comparados neste artigo é a forma como as características são extraídas. O artigo aborda *autoencoders* para a extração de características já os demais seguem a linha de *handcrafted features*, como também explicado em (RAHMAN et al., 2016) e (ROY et al., 2018).

Os resultados obtidos para o poço WELL00006201806181107 21 apresentaram métricas abaixo dos demais em todos modelos propostos. Esse poço apresenta um menor valor de dados normais em relação aos dados de falha (0,99%), em aproximadamente 124.000 observações apenas 1.000 destas são de dados da classe normal. Já em relação à média *F1 score* nos poços onde ocorreu o mesmo tipo de falha (falha 7), um resultado de 0,4922 para esse poço é encontrado enquanto que a média nos demais poços é de 0,8546. O poço WELL00006201708021230 00 também obteve resultados inferiores, quando comparados aos valores da média *F1 score* dentro de sua classe (falha 1), com valor de 0,5296 e a média dos demais em 0,9311. Estes valores foram obtidos dos modelos que apresentaram um melhor resultado, ou seja, dos detectores implementados com floresta de isolamento com redução para cem dimensões.

A partir dos resultados separados por classe de anomalia, observado na Tabela 4.4, onde são apresentadas as estimativas de tamanhos de janelas temporais (tempo de falha) normalmente utilizados para confirmar ocorrências de anomalias, conforme apresentado em Vargas (2019), pode ser observado que na falha 7 ocorre o pior resultado da média *F1*. Isso pode indicar que para a detecção de anomalia com uma janela de detecção (tempo de falha) maior, esta metodologia proposta pode não ser adequada e um número maior de atrasos deveria ser escolhido. Já para a Falha 2 que tem a menor janela temporal dos casos aplicados, o detector funcionou corretamente, obtendo a média métrica *F1 score* próxima a um.

Em comparação das duas técnicas de redução de dimensionalidade aplicadas nos dados de poços surgentes, os resultados foram distintos em relação ao número de cada redução proposta. Quando se comparada a redução de 500 para 100 variáveis o desempenho da técnica PCA para os *autoencoders* fica em torno de 96%. Já para a redução de 500 para 10 variáveis, esta

relação fica em torno de 91%. Nota-se que quanto menor é o valor da redução, mais próximos são os desempenhos das técnicas. Um fator que deve ser levado em consideração é o tempo de execução de cada abordagem, que são bem distintos. Para um poço que têm em torno de 100000 observações, a técnica PCA com redução de 500 variáveis para 100 o tempo de execução foi aproximadamente 30 vezes mais rápida. Sendo a duração de execução do *autoencoder* em torno de trinta minutos e a técnica PCA com execução em um minuto .

O modelo sem a inclusão de atrasos teve um desempenho inferior quando se comparados aos com atrasos e redução de dimensionalidade. Quanto mais elevada a quantidade dimensional dos dados melhor é o desempenho dos detectores, isto se deve a um maior conjunto de características disponíveis na construção dos modelos. Mas, quanto maior a quantidade de entradas, proporcionalmente irá crescer a complexidade no momento do treinamento dos modelos, demandando um tempo maior neste processo. A redução de dimensionalidade também auxilia no tratamento de dados com várias dimensões. A partir destas premissas a redução de dimensionalidade contribui para a construção de modelos para detecção de anomalias.

Tabela 4.4 – Resultados dados 3W de acordo com a classe de anomalia e o tempo de falha, no melhor modelo proposto.

Anomalia	Média *F1*	Tempo de falha
Falha 1	0,8505	12 h
Falha 2	0,9939	5 min -20 min
Falha 5	0,8023	12 h
Falha 6	0,8368	15 min
Falha 7	0,7330	72 h
Falha 8	0,7931	30 min - 5 h

Fonte: próprio autor.

5 CONCLUSÕES

Para os modelos desenvolvidos na detecção de falhas em um poço não-surgente, operado por *gas lift*, em sua grande maioria, apresentaram resultados satisfatórios, mesmo não havendo homogeneidade na quantidade de observações rotuladas de cada classe. Em especial, o modelo k-NN foi o que alcançou os melhores resultados, principalmente na métrica *recall*, considerada a mais importante para este tipo de problema.

A rede de *autoencoders* em cascata, utilizada em conjunto de outras técnicas para a classificação de falhas em poços de petróleo com elevação artificial por *gas lift*, apresenta uma grande aplicabilidade na diminuição da dimensionalidade dos dados. Os *autoencoders* possibilitaram a redução do tempo de treinamento, mantendo próximo o desempenho dos modelos ao comparar com modelos sem seu uso.

Os detectores de anomalias desenvolvidos neste trabalho, a partir dos dados do 3W *dataset*, que são são atribuídos a poços surgentes, apresentaram resultados superiores quando comparados com outros trabalhos da literatura, com uma diferença de até dezoito pontos percentuais para os modelos OCSVM e dez pontos percentuais para os modelos de Floresta de Isolamento. Isso demonstra que o uso de um número significativo de atrasos nas variáveis em conjunto com a técnica de *autoencoders* em cascata para a redução de dimensionalidade implica em uma melhor extração de características relevantes para este conjunto de dados.

A inclusão de atrasos faz com que o modelo se torne mais eficaz, obtendo mais informações sobre o processo a ser estudado. A partir dos resultados obtidos, foi observado que quanto maior a dimensão maior foi a média *F1 score*. O contraponto do aumento no número de atributos faz com que o conjunto de dados se expanda, tornando o desenvolvimento dos modelos mais vagaroso. A inclusão de *autoencoders* auxilia nesta tarefa dando mais agilidade ao treinamento. Esta técnica pode ser viável em abordagens de detecção de falhas em poços de petróleo *offshore*, disponibilizando informações mais precisas com a inclusão de características passadas em cada observação de uma série temporal.

Os detectores desenvolvidos aplicados à técnica de Floresta de Isolamento tiveram os melhores resultados em relação à média e ao desvio padrão, seguindo os demais resultados apresentados nos trabalhos já propostos para os dados 3W *dataset*.

No momento da concepção do projeto dos detectores também deve-se levar em consideração o tempo de execução no desenvolvimento dos modelos a serem implementados. As características dos processos que se deseja detectar anomalias, são fatores determinantes para

definir qual técnica de redução de dimensionalidade será empregada no processo de construção dos modelos. A estratégia dos *stacked autoencoders* conseguiu reduzir a dimensão do espaço latente e por isso o tempo de inferência foi reduzido.

Por fim, esta dissertação mostra que a utilização *autoencoders* em cascata pode ser viável a resolução nos mais variados tipos de problemas industriais, para os quais são necessárias respostas às anormalidades nos sistemas de monitoramento do processo de produção. Esse tratamento tende a reduzir a complexidade de ações diretas para retorno da normalidade nas plantas de processo.

5.1 Trabalhos Futuros

Para trabalhos futuros, recomenda-se a aplicação de outros modelos de detecção de falhas como outras técnicas de redução de características dos dados, verificando o seu desempenho em comparação às desenvolvidas. A otimização dos hiperparâmetros dos modelos tende a melhorar os resultados, assim como a construção de algoritmos que busquem os melhores pontos de atrasos das observações de cada série temporal.

A utilização dos modelos em outros poços de produção com elevação por *gas lift* permitirá, além de simulações e estudos da aplicabilidade em sistemas em tempo real, a verificação da generalização e compreensão dos resultados obtidos. Para os experimentos com poços não surgentes, avaliar resultados no nível de falhas em si, e não no nível de observações.

Para os experimentos com poços surgentes, realizar testes estatísticos adicionais para explicitar qual ou quais modelos geraram modelos melhores.

REFERÊNCIAS

ABDELLATIF, S. et al. A deep learning based on sparse auto-encoder with mcsa for broken rotor bar fault detection and diagnosis. In: **2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)**. [S.l.: s.n.], 2018. p. 1–6.

ABREU, P. E. O. G. B. **Detecção de Falhas em Sistemas Dinâmicos Sujeitos a Retardo no Tempo e Incertezas Paramétricas**. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte - MG, 7 2012.

AGGARWAL, C.; SATHE, S. An introduction to outlier ensembles. In: ____. [S.l.: s.n.], 2017. p. 1–34. ISBN 978-3-319-54764-0.

AGUIRRE, L. A. et al. Development of soft sensors for permanent downhole gauges in deepwater oil wells. **Control Engineering Practice**, v. 65, p. 83 – 99, 2017. ISSN 0967-0661. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0967066117301284>.

ALMEIDA, C. A. L. de. **Detecção de falhas em sistemas dinâmicos: abordagens imunoinspiradas**. Tese (Doutorado) — Universidade Federal de MInas Gerais, Belo Horizonte - MG, 2 2010.

ANP. **Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis 2019**. 2019. ed. Brasília - DF, 2019.

ANP. **Boletim da Produção de Petróleo e Gás Natural**. 126. ed. Brasília - DF, 2021.

ARAUJO, M.; AGUILAR, J.; APONTE, H. Fault detection system in gas lift well based on artificial immune system. In: **Proceedings of the International Joint Conference on Neural Networks, 2003.** [S.l.: s.n.], 2003. v. 3, p. 1673–1677 vol.3. ISSN 1098-7576.

BAGAJEWICZ, M. **Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations: Data Reconciliation, Gross Error Detection, and Instrumentation Upgrade**. McGraw-Hill Education, 2009. ISBN 9780071604727. Disponível em: <https://books.google.com.br/books?id=rXD1vBxmtQgC>.

BRAGA, A. de P.; CARVALHO, A. P. de Leon F. de; LUDERMIR, T. B. **Redes Neurais Artificiais - Teoria e Aplicações**. 2. ed. [S.l.]: LTC, 2007. ISBN 8521615647.

BRANDMAIER, A. et al. Structural equation model trees. **Psychological Methods**, v. 18, p. 71–86, 03 2013.

BRASIL, N. do. **Processamento de petróleo e gás**. Grupo Gen - LTC, 2011. ISBN 9788521619963. Disponível em: <https://books.google.com.br/books?id=l70YywAACAAJ>.

BREIMAN, L. et al. Classification and regression trees. belmont, ca: Wadsworth. **International Group**, v. 432, p. 151–166, 1984.

CAMPOS, M. et al. Anti-slug advanced control for offshore production platforms. In: . [S.l.: s.n.], 2015.

CAMPOS, M. et al. Advanced control systems for offshore production platforms. **Proceedings of the Annual Offshore Technology Conference**, v. 1, p. 141–153, 01 2013.

CARNEIRO, J. et al. Statistical characterization of two-phase slug flow in a horizontal pipe. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, v. 33, p. 251–258, 12 2010.

CASTRO, L. D.; ZUBEN, F. V. **Recent developments in biologically inspired computing**. Idea Group Pub., 2005. (E-Libro). ISBN 9781591403135. Disponível em: <https://books.google.com.br/books?id=t3lQAAAAMAAJ>.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 41, n. 3, jul. 2009. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/1541880.1541882>.

CHAUVIN, Y.; RUMELHART, D. E. (Ed.). **Backpropagation: Theory, Architectures, and Applications**. USA: L. Erlbaum Associates Inc., 1995. ISBN 0805812598.

DAI, X.; LIU, G.; LONG, Z. Discrete-time robust fault detection observer design: A genetic algorithm approach. In: **2008 7th World Congress on Intelligent Control and Automation**. [S.l.: s.n.], 2008. p. 2843–2848.

DEMADIS, K. D. et al. Industrial water systems: problems, challenges and solutions for the process industries. **Desalination**, v. 213, n. 1, p. 38 – 46, 2007. ISSN 0011-9164. New Water Culture of South East European Countries-AQUA 2005. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0011916407003001>.

DIEHL, F. C. et al. Oil production increase in unstable gas lift systems through nonlinear model predictive control. **Journal of Process Control**, v. 69, p. 58 – 69, 2018. ISSN 0959-1524. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0959152418301380>.

DUDA, R. O. et al. **Pattern Classification, 2nd Ed**. 2001.

EHLERS, R. S. Análise de séries temporais. **Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná**, v. 1, p. 1–118, 2007.

FAN, J.; WANG, W.; ZHANG, H. Autoencoder based high-dimensional data fault detection system. In: **2017 IEEE 15th International Conference on Industrial Informatics (INDIN)**. [S.l.: s.n.], 2017. p. 1001–1006.

FAWCETT, T. An introduction to roc analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861 – 874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Disponível em: <http://www.sciencedirect.com/science/article/pii/S016786550500303X>.

FILHO, H. dos S. R. **Otimização de gás lift na produção de petróleo: avaliação da curva de performance do poço**. 92 p. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, Rio de janeiro, 2011.

FOGLIATO, F.; RIBEIRO, J. **Confiabilidade e manutenção industrial**. Elsevier Editora Ltda., 2009. ISBN 9788535251883. Disponível em: <https://books.google.com.br/books?id=_GhSnuKRBtwC>.

GUERBAI, Y.; CHIBANI, Y.; HADJADJI, B. The effective use of the one-class svm classifier for reduced training samples and its application to handwritten signature verification. In: **2014 International Conference on Multimedia Computing and Systems (ICMCS)**. [S.l.: s.n.], 2014. p. 362–366.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference and prediction**. 2. ed. [S.l.]: Springer, 2009.

HAUSLER, R. H.; KRISHNAMURTHY, R. M.; SHERAR, B. W. A. Nace-2015-6147. In: ____. Dallas, Texas: NACE International, 2015. cap. Observation of Productivity Loss in Large Oil Wells due to Scale Formation without Apparent Production of Formation Brine, p. 10.

HAYKIN, S. **Redes Neurais Principio e Prática**. 2. ed. [S.l.]: bookmam, 2001. ISBN 9788573077186.

HE, K. et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: **2015 IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2015. p. 1026–1034. ISSN 2380-7504.

ISERMANN, R.; BALLé, P. Trends in the application of model-based fault detection and diagnosis of technical processes. **Control Engineering Practice**, v. 5, n. 5, p. 709 – 719, 1997. ISSN 0967-0661. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0967066197000531>.

IZENMAN, A. J. Linear discriminant analysis. In: ____. **Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning**. New York, NY: Springer New York, 2008. p. 237–280. ISBN 978-0-387-78189-1. Disponível em: <https://doi.org/10.1007/978-0-387-78189-1_8>.

JUNIOR, W. F. et al. Detecção de anomalias em poços produtores de petróleo usando aprendizado de máquina. In: **Congresso Brasileiro de Automática (CBA)**. [S.l.: s.n.], 2020. v. 2, n. 1.

KADHIM, A. Survey on supervised machine learning techniques for automatic text classification. **Artificial Intelligence Review**, v. 52, 06 2019.

KAPLAN, A.; HAENLEIN, M. Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons**, v. 62, n. 1, p. 15 – 25, 2019. ISSN 0007-6813. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0007681318301393>.

KHATTREE, R.; NAIK, D. N. **Multivariate Data Reduction and Discrimination with SAS Software**. 1st. ed. [S.l.]: SAS Publishing, 2000. ISBN 1580256961.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: . [S.l.: s.n.], 1995. v. 14.

KRAWCZYK, B. et al. Ensemble learning for data stream analysis: A survey. **Information Fusion**, v. 37, p. 132 – 156, 2017. ISSN 1566-2535. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1566253516302329>.

LEEMIS, L. **Reliability: Probabilistic Models and Statistical Methods**. Prentice Hall, 1995. (Prentice-Hall international series in industrial and systems engineering). ISBN 9780137205172. Disponível em: <https://books.google.com.br/books?id=2Z6XQgAACAAJ>.

LIDEN, R. **Algoritmos Genéticos (2a edição)**. BRASPORT, 2008. ISBN 9788574523736. Disponível em: <https://books.google.com.br/books?id=it0kv6UsEMEC>.

LIU, F. T.; TING, K. M.; ZHOU, Z. Isolation forest. In: **2008 Eighth IEEE International Conference on Data Mining**. [S.l.: s.n.], 2008. p. 413–422.

LU, C. et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. **Signal Processing**, v. 130, 07 2016.

LUCCHESI, C. F. Petróleo. **Estudos Avançados**, scielo, v. 12, p. 17 – 40, 08 1998. ISSN 0103-4014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40141998000200003&nrm=iso>.

LUGER, G. **Inteligência Artificial: Estruturas e estratégias para a solução de problemas complexos**. [S.l.]: Bookman, 2004. ISBN 9788536303963.

MACHADO, F. **Big Data O Futuro dos Dados e Aplicações**. Editora Saraiva, 2018. ISBN 9788536527611. Disponível em: <https://books.google.com.br/books?id=2LdiDwAAQBAJ>.

MAGLARAS, L.; JIANG, J. Intrusion detection in scada systems using machine learning techniques. In: . [S.l.: s.n.], 2014.

MEGLIO, F. D. 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. ISSN 0959-1524. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0959152412000637>.

MORÉ, J. J. The levenberg-marquardt algorithm: Implementation and theory. In: WATSON, G. (Ed.). **Numerical Analysis**. [S.l.]: Springer Berlin Heidelberg, 1978, (Lecture Notes in Mathematics, v. 630). p. 105–116.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. [S.l.]: Blucher, 2006. ISBN 9788521213512.

MUCHERINO, A.; PAPAJORGJI, P. J.; PARDALOS, P. M. k-nearest neighbor classification. In: ____. **Data Mining in Agriculture**. New York, NY: Springer New York, 2009. p. 83–106. ISBN 978-0-387-88615-2. Disponível em: <https://doi.org/10.1007/978-0-387-88615-2_4>.

NASCIMENTO, R. S. F. do et al. Detecção de falhas com stacked autoencoders e técnicas de reconhecimento de padrões em poços de petróleo operados por gas lift. In: **Congresso Brasileiro de Automática (CBA)**. [S.l.: s.n.], 2020. v. 2, n. 1.

NAVES, R. **Um estudo de reconhecimento de sons pulmonares baseado em técnicas de inteligência computacional**. 94 p. Dissertação (Mestrado) — Universidade Federal de Lavras, lavras, 2015.

NGUYEN, A. N. of L. Tutorial on support vector machine. **Special Issue "Some Novel Algorithms for Global Optimization and Relevant Subjects", Applied and Computational Mathematics (ACM)**, v. 6, p. 1–15, 06 2016.

NILCHIANI, R.; EDWARDS, C. M.; GANGULY, A. Introducing a tipping point measure in explaining disruptive technology. In: **2019 International Symposium on Systems Engineering (ISSE)**. [S.l.: s.n.], 2019. p. 1–5.

OLIVEIRA, R.; GONZALEZ, G.; SANTIAGO, V. Efeito do campo magnético na precipitação de parafinas. **Química Nova**, v. 21, p. 11 – 15, 02 1998.

PARK, P. et al. Fault detection and diagnosis using combined autoencoder and long short-term memory network. **Sensors**, v. 19, n. 21, 2019. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/19/21/4612>.

PEYERL, D. Front matter. In: ____. **O petróleo no Brasil: exploração, capacitação técnica e ensino de geociências (1864-1968)**. Dgo - digital original. SciELO – Editora UFABC, 2017. p. I–XVI. ISBN 9788568576588. Disponível em: <http://www.jstor.org/stable/10.7476/9788568576786.1>.

PLUCENIO, A. **Automação da produção de poços de petróleo operando com elevação artificial por injeção contínua de gás**. 118 p. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, Florianópolis, 2003.

RAHMAN, A. et al. A comparison of autoencoder and statistical features for cattle behaviour classification. In: **2016 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2016. p. 2954–2960.

RICHMAN, J. S. Chapter thirteen - multivariate neighborhood sample entropy: A method for data reduction and prediction of complex data. In: JOHNSON, M. L.; BRAND, L. (Ed.). **Computer Methods, Part C**. [S.l.]: Academic Press, 2011, (Methods in Enzymology, v. 487). p. 397 – 408.

RODRIGUEZ, J. D.; PEREZ, A.; LOZANO, J. A. Sensitivity analysis of k-fold cross validation in prediction error estimation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, n. 3, p. 569–575, 2010.

ROY, M. et al. A stacked autoencoder neural network based automated feature extraction method for anomaly detection in on-line condition monitoring. In: **2018 IEEE Symposium Series on Computational Intelligence (SSCI)**. [S.l.: s.n.], 2018. p. 1501–1507.

SANTOS, I. H. et al. Hydrate failure detection in production and injection lines using model and data-driven approaches. In: **Rio Oil Gas Expo and Conference 2018**. Rio de Janeiro - RJ: [s.n.], 2018.

SCHIAVI, M. T.; HOFFMANN, W. A. M. Cenário petrolífero: sua evolução, principais produtores e tecnologias. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 13, n. 2, p. 259–278, maio 2015.

SCHLAG, S.; SCHMITT, M.; SCHULZ, C. Faster support vector machines. In: ____. [S.l.: s.n.], 2019. p. 199–210. ISBN 978-1-61197-549-9.

SCHöLKOPF, B. et al. Estimating the Support of a High-Dimensional Distribution. **Neural Computation**, v. 13, n. 7, p. 1443–1471, 07 2001. ISSN 0899-7667. Disponível em: <https://doi.org/10.1162/089976601750264965>.

SHARMA, V.; RAI, S.; DEV, A. A comprehensive study of artificial neural networks. **International Journal of Advanced Research in Computer Science and Software Engineering**, v. 2, n. 2, p. 279–284, 10 2012.

SHIN, H. et al. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 8, p. 1930–1943, 2013.

SILVA, I. J. F. lima Verçosa; Vitória Camila Paixão dos S. F. M. S. . M. A. L. S. J. S. D. Formação de hidratos em perfurações de poços em águas profundas e ultra profundas. **III CONEPETRO Congresso Nacional de Engenharia de Patróleo, Gás Natural e Biocombustíveis**, n. 3, 8 2018.

SIMON, P. **Too Big to Ignore: The Business Case for Big Data**. Wiley, 2013. (Wiley and SAS Business Series). ISBN 9781118642108. Disponível em: <https://books.google.com.br/books?id=Dn-Gdoh66sgC>.

TAITEL, Y.; BARNEA, D. Two-phase slug flow. In: HARTNETT, J. P.; IRVINE, T. F. (Ed.). Elsevier, 1990, (Advances in Heat Transfer, v. 20). p. 83 – 132. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0065271708700261>.

THEYAB, M. Severe slugging control: Simulation of real case study. v. 2, 03 2018.

THOMAS, J. **Fundamentos de engenharia de petróleo**. Interciência, 2004. ISBN 9788571930995. Disponível em: <https://books.google.com.br/books?id=yKyqPgAACAAJ>.

TING, K. M. Confusion matrix. In: ____. **Encyclopedia of Machine Learning and Data Mining**. Boston, MA: Springer US, 2017. p. 260–260. ISBN 978-1-4899-7687-1. Disponível em: <https://doi.org/10.1007/978-1-4899-7687-1_50>.

VARGAS, R. E. V. **Base de Dados e Benchmarks para Prognóstico de Anomalias em Sistemas de Elevação de Petróleo**. Tese (Doutorado) — Universidade Federal do Espírito Santo, Vitória - ES, 8 2019.

VARGAS, R. E. V. 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. ISSN 0920-4105. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0920410519306357>.

WEN, L.; GAO, L.; LI, X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, v. 49, n. 1, p. 136–144, 2019.

WYLD, D. et al. **Advances in Computing and Information Technology: First International Conference, ACITY 2011, Chennai, India, July 15-17, 2011, Proceedings**. Springer Berlin Heidelberg, 2011. (Communications in Computer and Information Science). ISBN 9783642225550. Disponível em: <https://books.google.com.br/books?id=JHgQBwAAQBAJ>.

YU, J.; ZHANG, C. Manifold regularized stacked autoencoders-based feature learning for fault detection in industrial processes. **Journal of Process Control**, v. 92, p. 119 – 136, 2020. ISSN 0959-1524. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0959152420302353>.

APÊNDICE A – Tabelas com os Resultados dos Detectores para os Poços Surgentes

Tabela 1 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precission* e *F1 score*, com 25% dos dados normais e duas dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	0,9833	1,000	0,9916
WELL00002201401262000 50	0,9786	0,9987	0,9886
WELL00006201708010636 14	0,8189	1,000	0,9004
WELL00006201708021230 00	0,3890	0,5544	0,4572
WELL00006201806180602 45	1,000	0,9611	0,9801
WELL00002201311040141 01	1,000	0,9923	0,9962
WELL00009201703131608 04	1,000	0,9910	0,9955
WELL00010201712182001 31	1,000	0,9680	0,9837
WELL00015201706201603 49	1,000	0,8944	0,9443
WELL00015201710131400 47	0,6597	0,8960	0,7599
WELL00016201804050203 45	0,9520	0,6815	0,7944
WELL00016201804261420 05	1,000	0,5210	0,6850
WELL00016201804261451 08	0,6347	0,8230	0,7167
WELL00016201805172223 22	1,000	0,6563	0,7925
WELL00017201403171517 43	0,7473	0,8786	0,8077
WELL00017201403180231 41	0,3712	0,8787	0,5219
WELL00017201403181602 20	0,4208	0,9710	0,5872
WELL00017201403190404 53	0,7953	0,956	0,8683
WELL00017201403191414 50	1,000	0,6393	0,7799
WELL00002201402121703 33	0,9976	0,6814	0,8097
WELL00002201403011517 00	0,9989	0,6188	0,7642
WELL00002201403251703 04	0,8159	1,000	0,8986
WELL00004201710311815 09	0,9871	0,7267	0,8371
WELL00004201710311930 25	1,000	0,5232	0,6870
WELL00004201710312000 59	1,000	0,5018	0,6683
WELL00001201702262203 09	0,4761	0,9406	0,6322
WELL00006201806181107 21	0,4364	0,2894	0,3480
WELL00006201806201813 48	0,9822	0,7314	0,8385
WELL00021201705090135 17	0,9871	0,9674	0,9771
WELL00020201204101923 26	0,3992	0,9694	0,5655
WELL00019201703011823 17	0,7941	0,8154	0,8046

Fonte: própio autor.

Tabela 2 – Resultados dados 3W para o modelo one-class SVM, *recall*, *precision* e *F1 score*, com 25% dos dados normais e duas dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,7121	0,8318
WELL00002201401262000 50	0,9811	1,000	0,9905
WELL00006201708010636 14	1,000	0,6441	0,7836
WELL00006201708021230 00	0,1396	0,2703	0,1841
WELL00006201806180602 45	0,9548	1,000	0,9769
WELL00002201311040141 01	0,9885	1,000	0,9942
WELL00009201703131608 04	1,000	0,9258	0,9614
WELL00010201712182001 31	0,9724	1,000	0,9860
WELL00015201706201603 49	0,9297	0,1453	0,2514
WELL00015201710131400 47	0,9361	0,4250	0,5846
WELL00016201804050203 45	1,000	0,3083	0,4713
WELL00016201804261420 05	1,000	0,2934	0,4537
WELL00016201804261451 08	0,6023	0,5072	0,5506
WELL00016201805172223 22	1,000	0,4666	0,6363
WELL00017201403171517 43	0,5341	0,6810	0,5987
WELL00017201403180231 41	0,4150	0,4037	0,4093
WELL00017201403181602 20	0,5107	0,5156	0,5131
WELL00017201403190404 53	0,7945	0,5311	0,6366
WELL00017201403191414 50	1,000	0,4578	0,6280
WELL00002201402121703 33	1,000	0,5358	0,6978
WELL00002201403011517 00	0,9931	0,4334	0,6034
WELL00002201403251703 04	0,8199	1,000	0,9011
WELL00004201710311815 09	0,9929	0,6609	0,7935
WELL00004201710311930 25	1,000	0,2278	0,3711
WELL00004201710312000 59	1,000	0,3054	0,4679
WELL00001201702262203 09	0,6523	0,5007	0,5665
WELL00006201806181107 21	0,1300	0,3970	0,1958
WELL00006201806201813 48	0,7966	0,6477	0,7145
WELL00021201705090135 17	0,8113	0,6790	0,7393
WELL00020201204101923 26	1,000	0,9477	0,9731
WELL00019201703011823 17	0,9176	0,8571	0,8864

Fonte: própio autor.

Tabela 3 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precision* e *F1 score*, com 60% dos dados normais e duas dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	0,9801	1,000	0,9900
WELL00002201401262000 50	0,9700	0,8476	0,9047
WELL00006201708010636 14	0,9823	1,000	0,9911
WELL00006201708021230 00	0,3905	0,6634	0,4916
WELL00006201806180602 45	1,000	0,9995	0,9998
WELL00002201311040141 01	1,000	0,9993	0,9997
WELL00009201703131608 04	1,000	0,9993	0,9996
WELL00010201712182001 31	1,000	0,9989	0,9994
WELL00015201706201603 49	0,8681	1,000	0,9294
WELL00015201710131400 47	0,6738	0,9284	0,7808
WELL00016201804050203 45	0,9520	0,6815	0,7944
WELL00016201804261420 05	1,000	0,5960	0,7469
WELL00016201804261451 08	0,6347	0,823	0,7167
WELL00016201805172223 22	0,9322	1,000	0,9649
WELL00017201403171517 43	0,5476	0,9524	0,6954
WELL00017201403180231 41	0,5062	0,8474	0,6338
WELL00017201403181602 20	0,5132	0,7970	0,6244
WELL00017201403190404 53	0,6885	0,9269	0,7908
WELL00017201403191414 50	0,7785	0,9707	0,864
WELL00002201402121703 33	0,9343	0,7871	0,8544
WELL00002201403011517 00	0,9945	0,6522	0,7878
WELL00002201403251703 04	0,7200	0,9921	0,8344
WELL00004201710311815 09	0,9636	0,8025	0,8757
WELL00004201710311930 25	1,000	0,5906	0,7426
WELL00004201710312000 59	1,000	0,5819	0,7357
WELL00001201702262203 09	0,5338	0,9031	0,671
WELL00006201806181107 21	0,3750	0,4819	0,4218
WELL00006201806201813 48	0,9686	0,7801	0,8642
WELL00021201705090135 17	0,9529	0,9611	0,957
WELL00020201204101923 26	0,5856	0,8442	0,6915
WELL00019201703011823 17	0,6743	0,7975	0,7307

Fonte: própio autor.

Tabela 4 – Resultados dados 3W para o modelo one-class SVM, *recall*, *precision* e *F1 score*, com 60% dos dados normais duas dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,7854	0,8779
WELL00002201401262000 50	0,9952	0,9852	0,9901
WELL00006201708010636 14	1,000	0,6732	0,8046
WELL00006201708021230 00	0,1835	0,2964	0,2266
WELL00006201806180602 45	0,9257	0,9854	0,9546
WELL00002201311040141 01	0,9714	0,9934	0,9822
WELL00009201703131608 04	1,000	0,9467	0,9726
WELL00010201712182001 31	0,9874	1,000	0,9936
WELL00015201706201603 49	0,9451	0,2415	0,3846
WELL00015201710131400 47	0,8922	0,4105	0,5622
WELL00016201804050203 45	1,000	0,4254	0,5969
WELL00016201804261420 05	1,000	0,3321	0,4986
WELL00016201804261451 08	0,5841	0,4875	0,5314
WELL00016201805172223 22	0,9845	0,4836	0,6486
WELL00017201403171517 43	0,6020	0,7054	0,6496
WELL00017201403180231 41	0,4423	0,5019	0,4702
WELL00017201403181602 20	0,5107	0,5156	0,5131
WELL00017201403190404 53	0,7707	0,5796	0,6616
WELL00017201403191414 50	1,000	0,4954	0,6626
WELL00002201402121703 33	0,9854	0,5158	0,6772
WELL00002201403011517 00	0,9831	0,4412	0,6091
WELL00002201403251703 04	0,8399	1,000	0,9130
WELL00004201710311815 09	0,979	0,6805	0,8029
WELL00004201710311930 25	0,9965	0,3154	0,4791
WELL00004201710312000 59	1,000	0,2987	0,4600
WELL00001201702262203 09	0,7066	0,5114	0,5934
WELL00006201806181107 21	0,2099	0,3847	0,2716
WELL00006201806201813 48	0,7865	0,6855	0,7325
WELL00021201705090135 17	0,8224	0,6988	0,7556
WELL00020201204101923 26	1,000	0,9331	0,9654
WELL00019201703011823 17	0,9122	0,8547	0,8825

Fonte: própio autor.

Tabela 5 – Resultados dados 3W para o modelo one-class SVM, *recall*, *precision* e *F1 score*, com 60% dos dados normais e cem dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,9368	0,9674
WELL00002201401262000 50	1,000	0,9244	0,9607
WELL00006201708010636 14	0,9854	0,6631	0,7927
WELL00006201708021230 00	0,2506	0,3604	0,2956
WELL00006201806180602 45	0,9521	0,9354	0,9437
WELL00002201311040141 01	0,9822	0,9855	0,9838
WELL00009201703131608 04	1,000	0,9668	0,9831
WELL00010201712182001 31	0,9924	1,000	0,9962
WELL00015201706201603 49	0,9647	0,2822	0,4367
WELL00015201710131400 47	0,9245	0,4815	0,6332
WELL00016201804050203 45	1,000	0,5395	0,7009
WELL00016201804261420 05	1,000	0,3702	0,5404
WELL00016201804261451 08	0,6027	0,5148	0,5553
WELL00016201805172223 22	0,9914	0,4900	0,6558
WELL00017201403171517 43	0,6450	0,7387	0,6887
WELL00017201403180231 41	0,5058	0,5471	0,5256
WELL00017201403181602 20	0,5378	0,5498	0,5437
WELL00017201403190404 53	0,8041	0,6035	0,6895
WELL00017201403191414 50	1,000	0,5703	0,7264
WELL00002201402121703 33	0,9732	0,6024	0,7442
WELL00002201403011517 00	0,9682	0,5324	0,6870
WELL00002201403251703 04	0,8697	1,000	0,9303
WELL00004201710311815 09	0,9632	0,7532	0,8454
WELL00004201710311930 25	1,000	0,3952	0,5665
WELL00004201710312000 59	1,000	0,3647	0,5345
WELL00001201702262203 09	0,7458	0,5714	0,6471
WELL00006201806181107 21	0,2849	0,4125	0,3370
WELL00006201806201813 48	0,8498	0,7154	0,7768
WELL00021201705090135 17	0,862	0,6755	0,7574
WELL00020201204101923 26	1,000	0,9587	0,9789
WELL00019201703011823 17	0,9304	0,8688	0,8985

Fonte: própio autor.

Tabela 6 – Resultados dados 3W para o modelo one-class SVM, *recall*, *precision* e *F1 score*, com 60% dos dados normais e cinquenta dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,9159	0,9561
WELL00002201401262000 50	1,000	0,9114	0,9536
WELL00006201708010636 14	0,9458	0,6430	0,7655
WELL00006201708021230 00	0,1835	0,2964	0,2267
WELL00006201806180602 45	0,9443	0,9952	0,9691
WELL00002201311040141 01	0,9810	0,9741	0,9775
WELL00009201703131608 04	0,9921	0,9574	0,9744
WELL00010201712182001 31	0,9755	1,000	0,9876
WELL00015201706201603 49	0,9608	0,2721	0,4241
WELL00015201710131400 47	0,9178	0,4366	0,5917
WELL00016201804050203 45	1,000	0,4638	0,6337
WELL00016201804261420 05	1,000	0,3391	0,5065
WELL00016201804261451 08	0,5937	0,5019	0,5440
WELL00016201805172223 22	0,9896	0,5036	0,6675
WELL00017201403171517 43	0,6359	0,7126	0,6721
WELL00017201403180231 41	0,4539	0,5214	0,4853
WELL00017201403181602 20	0,5304	0,5290	0,5297
WELL00017201403190404 53	0,7918	0,5843	0,6724
WELL00017201403191414 50	1,000	0,5234	0,6871
WELL00002201402121703 33	0,9899	0,5824	0,7333
WELL00002201403011517 00	0,9633	0,5029	0,6608
WELL00002201403251703 04	0,8585	1,000	0,9239
WELL00004201710311815 09	0,9524	0,7412	0,8336
WELL00004201710311930 25	1,000	0,3841	0,5550
WELL00004201710312000 59	1,000	0,3429	0,5107
WELL00001201702262203 09	0,7466	0,5770	0,6509
WELL00006201806181107 21	0,2763	0,4008	0,3271
WELL00006201806201813 48	0,8498	0,7198	0,7687
WELL00021201705090135 17	0,8198	0,6900	0,7493
WELL00020201204101923 26	1,000	0,9355	0,9667
WELL00019201703011823 17	0,9254	0,8730	0,8984

Fonte: própio autor.

Tabela 7 – Resultados dados 3W para o modelo one-class SVM, *recall*, *precision* e *F1 score*, com 60% dos dados normais e dez dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,8704	0,9307
WELL00002201401262000 50	1,000	0,8865	0,9398
WELL00006201708010636 14	0,9566	0,6148	0,7485
WELL00006201708021230 00	0,1736	0,2855	0,2159
WELL00006201806180602 45	0,9257	0,9791	0,9517
WELL00002201311040141 01	0,9801	0,9899	0,9850
WELL00009201703131608 04	1,000	0,9710	0,9853
WELL00010201712182001 31	0,9648	1,000	0,9821
WELL00015201706201603 49	0,9488	0,2597	0,4078
WELL00015201710131400 47	0,9004	0,4306	0,5826
WELL00016201804050203 45	1,000	0,4394	0,6105
WELL00016201804261420 05	1,000	0,3234	0,4887
WELL00016201804261451 08	0,5894	0,4933	0,5371
WELL00016201805172223 22	0,9833	0,4958	0,6592
WELL00017201403171517 43	0,6277	0,7185	0,6700
WELL00017201403180231 41	0,4499	0,5109	0,4785
WELL00017201403181602 20	0,5387	0,5164	0,5273
WELL00017201403190404 53	0,7527	0,5857	0,6588
WELL00017201403191414 50	1,000	0,5584	0,7166
WELL00002201402121703 33	0,9652	0,5436	0,6955
WELL00002201403011517 00	0,9341	0,4720	0,6271
WELL00002201403251703 04	0,8352	1,000	0,9102
WELL00004201710311815 09	0,9493	0,7123	0,8139
WELL00004201710311930 25	0,9821	0,3602	0,5271
WELL00004201710312000 59	1,000	0,3357	0,5027
WELL00001201702262203 09	0,7351	0,5299	0,6159
WELL00006201806181107 21	0,2687	0,3991	0,3212
WELL00006201806201813 48	0,7927	0,7154	0,7521
WELL00021201705090135 17	0,8257	0,7360	0,7783
WELL00020201204101923 26	0,9985	0,9211	0,9583
WELL00019201703011823 17	0,9214	0,8680	0,8939

Fonte: própio autor.

Tabela 8 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precision* e *F1 score*, com 60% dos dados normais e cem dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,9759	0,9878
WELL00002201401262000 50	1,000	0,9951	0,9975
WELL00006201708010636 14	0,9452	0,6142	0,7446
WELL00006201708021230 00	0,4255	0,7012	0,5296
WELL00006201806180602 45	1,000	0,9887	0,9943
WELL00002201311040141 01	0,9924	0,9837	0,9880
WELL00009201703131608 04	1,000	0,9951	0,9975
WELL00010201712182001 31	1,000	0,9925	0,9962
WELL00015201706201603 49	0,9241	1,000	0,9606
WELL00015201710131400 47	0,7015	0,9433	0,8046
WELL00016201804050203 45	0,9628	0,7306	0,8308
WELL00016201804261420 05	1,000	0,6438	0,7833
WELL00016201804261451 08	0,6962	0,8325	0,7583
WELL00016201805172223 22	0,9523	1,000	0,9756
WELL00017201403171517 43	0,5943	0,9658	0,7358
WELL00017201403180231 41	0,5309	0,8667	0,6585
WELL00017201403181602 20	0,5365	0,8143	0,6468
WELL00017201403190404 53	0,7168	0,9367	0,8121
WELL00017201403191414 50	0,7625	0,9826	0,8587
WELL00002201402121703 33	0,9428	0,8447	0,8911
WELL00002201403011517 00	1,000	0,6702	0,8025
WELL00002201403251703 04	0,7601	0,9821	0,8570
WELL00004201710311815 09	0,9822	0,8400	0,9056
WELL00004201710311930 25	1,000	0,6542	0,7910
WELL00004201710312000 59	1,000	0,6306	0,7735
WELL00001201702262203 09	0,6544	0,9863	0,7868
WELL00006201806181107 21	0,4487	0,5450	0,4922
WELL00006201806201813 48	0,9807	0,8665	0,9201
WELL00021201705090135 17	0,9744	0,9122	0,9423
WELL00020201204101923 26	0,6004	0,8366	0,6991
WELL00019201703011823 17	0,6833	0,8022	0,7380

Fonte: própio autor.

Tabela 9 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precision* e *F1 score*, com 60% dos dados normais e cinquenta dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,9541	0,9765
WELL00002201401262000 50	1,000	0,9335	0,9656
WELL00006201708010636 14	0,9721	0,6357	0,7687
WELL00006201708021230 00	0,4721	0,5988	0,5280
WELL00006201806180602 45	1,000	0,9956	0,9978
WELL00002201311040141 01	0,9957	0,9833	0,9895
WELL00009201703131608 04	1,000	0,9867	0,9933
WELL00010201712182001 31	1,000	0,9822	0,9910
WELL00015201706201603 49	0,9006	1,000	0,9477
WELL00015201710131400 47	0,6993	0,9301	0,7984
WELL00016201804050203 45	0,9577	0,7036	0,8112
WELL00016201804261420 05	1,000	0,6357	0,7773
WELL00016201804261451 08	0,6763	0,8306	0,7456
WELL00016201805172223 22	0,9306	1,000	0,9641
WELL00017201403171517 43	0,5821	0,9544	0,7231
WELL00017201403180231 41	0,5259	0,8587	0,6523
WELL00017201403181602 20	0,5235	0,8022	0,6336
WELL00017201403190404 53	0,7085	0,9337	0,8057
WELL00017201403191414 50	0,7785	0,9707	0,8640
WELL00002201402121703 33	0,9402	0,8233	0,8779
WELL00002201403011517 00	1,000	0,6604	0,7955
WELL00002201403251703 04	0,7400	0,9803	0,8434
WELL00004201710311815 09	0,9807	0,8378	0,9036
WELL00004201710311930 25	1,000	0,6371	0,7783
WELL00004201710312000 59	1,000	0,6143	0,7611
WELL00001201702262203 09	0,6017	0,9266	0,7296
WELL00006201806181107 21	0,4098	0,4720	0,4387
WELL00006201806201813 48	0,9773	0,8004	0,8800
WELL00021201705090135 17	0,9370	0,9408	0,9389
WELL00020201204101923 26	0,6077	0,8752	0,7173
WELL00019201703011823 17	0,6519	0,7570	0,7005

Fonte: própio autor.

Tabela 10 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precision* e *F1 score*, com 60% dos dados normais e dez dimensões.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,9432	0,9708
WELL00002201401262000 50	1,000	0,9753	0,9875
WELL00006201708010636 14	0,9256	0,5857	0,7174
WELL00006201708021230 00	0,3727	0,5421	0,4417
WELL00006201806180602 45	1,000	0,9896	0,9948
WELL00002201311040141 01	0,9975	0,9544	0,9755
WELL00009201703131608 04	1,000	0,9708	0,9852
WELL00010201712182001 31	1,000	0,9713	0,9854
WELL00015201706201603 49	0,9025	1,000	0,9488
WELL00015201710131400 47	0,6809	0,9406	0,7900
WELL00016201804050203 45	0,9536	0,7306	0,8273
WELL00016201804261420 05	1,000	0,6139	0,7608
WELL00016201804261451 08	0,6568	0,8015	0,7220
WELL00016201805172223 22	0,9282	1,000	0,9628
WELL00017201403171517 43	0,607	0,8803	0,7185
WELL00017201403180231 41	0,5300	0,8530	0,6538
WELL00017201403181602 20	0,5506	0,7964	0,6511
WELL00017201403190404 53	0,6957	0,9138	0,7900
WELL00017201403191414 50	0,7885	0,9434	0,8590
WELL00002201402121703 33	0,9002	0,8112	0,8534
WELL00002201403011517 00	0,9631	0,6808	0,7977
WELL00002201403251703 04	0,7634	0,9425	0,8435
WELL00004201710311815 09	0,9488	0,8025	0,8695
WELL00004201710311930 25	1,000	0,6238	0,7683
WELL00004201710312000 59	1,000	0,6128	0,7599
WELL00001201702262203 09	0,5668	0,9194	0,7013
WELL00006201806181107 21	0,4155	0,4947	0,4517
WELL00006201806201813 48	0,9706	0,7725	0,8603
WELL00021201705090135 17	0,9529	0,9511	0,9520
WELL00020201204101923 26	0,5455	0,8787	0,6731
WELL00019201703011823 17	0,6957	0,7831	0,7368

Fonte: própio autor.

Tabela 11 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precision* e *F1 score*, com 60% dos dados normais e cem dimensões com PCA.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	0,9736	0,9432	0,9582
WELL00002201401262000 50	0,8644	0,9533	0,9067
WELL00006201708010636 14	0,9013	0,6257	0,7386
WELL00006201708021230 00	0,4236	0,5975	0,4957
WELL00006201806180602 45	0,8934	0,8874	0,8904
WELL00002201311040141 01	0,9831	0,9657	0,9743
WELL00009201703131608 04	1,000	0,9724	0,9860
WELL00010201712182001 31	1,000	0,9852	0,9925
WELL00015201706201603 49	0,9158	0,9624	0,9385
WELL00015201710131400 47	0,7009	0,9133	0,7931
WELL00016201804050203 45	0,8263	0,8087	0,8174
WELL00016201804261420 05	0,9252	0,6156	0,7393
WELL00016201804261451 08	0,5133	0,6041	0,5550
WELL00016201805172223 22	0,7557	0,9229	0,8310
WELL00017201403171517 43	0,5844	0,9557	0,7253
WELL00017201403180231 41	0,4962	0,7788	0,6062
WELL00017201403181602 20	0,5260	0,6451	0,5795
WELL00017201403190404 53	0,7536	0,8627	0,8045
WELL00017201403191414 50	0,7045	0,9324	0,8026
WELL00002201402121703 33	0,9287	0,8378	0,8809
WELL00002201403011517 00	0,8563	0,8136	0,8344
WELL00002201403251703 04	0,7456	0,8858	0,8097
WELL00004201710311815 09	0,9334	0,8274	0,8772
WELL00004201710311930 25	0,9254	0,7065	0,8013
WELL00004201710312000 59	0,9539	0,6688	0,7863
WELL00001201702262203 09	0,7808	0,9034	0,8376
WELL00006201806181107 21	0,4068	0,4387	0,4221
WELL00006201806201813 48	0,8757	0,8471	0,8612
WELL00021201705090135 17	0,9469	0,9557	0,9513
WELL00020201204101923 26	0,6258	0,8674	0,7271
WELL00019201703011823 17	0,6935	0,7939	0,7403

Fonte: própio autor.

Tabela 12 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precission* e *F1 score*, com 60% dos dados normais e dez dimensões com PCA.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	0,9562	0,9154	0,9354
WELL00002201401262000 50	0,8374	0,8454	0,8414
WELL00006201708010636 14	0,4722	0,6971	0,5630
WELL00006201708021230 00	0,3963	0,4631	0,4271
WELL00006201806180602 45	0,9687	0,8833	0,9240
WELL00002201311040141 01	0,9468	0,9248	0,9357
WELL00009201703131608 04	0,9962	0,9851	0,9906
WELL00010201712182001 31	0,9665	0,9767	0,9716
WELL00015201706201603 49	0,8136	0,9411	0,8727
WELL00015201710131400 47	0,5559	0,8136	0,6605
WELL00016201804050203 45	0,8532	0,6825	0,7584
WELL00016201804261420 05	0,7156	0,5221	0,6037
WELL00016201804261451 08	0,6558	0,8262	0,7312
WELL00016201805172223 22	0,8336	0,9106	0,8704
WELL00017201403171517 43	0,5216	0,7436	0,6131
WELL00017201403180231 41	0,4962	0,7788	0,6062
WELL00017201403181602 20	0,5032	0,7460	0,6010
WELL00017201403190404 53	0,6115	0,7534	0,6751
WELL00017201403191414 50	0,7254	0,8158	0,7679
WELL00002201402121703 33	0,9102	0,6419	0,7529
WELL00002201403011517 00	0,9048	0,6359	0,7469
WELL00002201403251703 04	0,8934	0,9525	0,9220
WELL00004201710311815 09	0,9169	0,7269	0,8109
WELL00004201710311930 25	0,9363	0,5324	0,6788
WELL00004201710312000 59	0,9278	0,5712	0,7071
WELL00001201702262203 09	0,4436	0,7269	0,5510
WELL00006201806181107 21	0,4087	0,4705	0,4374
WELL00006201806201813 48	0,8618	0,7025	0,7740
WELL00021201705090135 17	0,8540	0,6455	0,7353
WELL00020201204101923 26	0,4863	0,8128	0,6085
WELL00019201703011823 17	0,6015	0,6968	0,6457

Fonte: própio autor.

Tabela 13 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precission* e *F1 score* sem atrsos.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	0,8465	0,8322	0,8393
WELL00002201401262000 50	0,7769	0,7684	0,7726
WELL00006201708010636 14	0,4452	0,6527	0,5293
WELL00006201708021230 00	0,4568	0,3814	0,4157
WELL00006201806180602 45	0,7539	0,7275	0,7405
WELL00002201311040141 01	0,9269	0,9566	0,9415
WELL00009201703131608 04	0,9164	0,9437	0,9298
WELL00010201712182001 31	0,9252	0,9604	0,9425
WELL00015201706201603 49	0,6724	0,8173	0,7378
WELL00015201710131400 47	0,6117	0,7801	0,6857
WELL00016201804050203 45	0,7281	0,8008	0,7627
WELL00016201804261420 05	0,6224	0,4712	0,5363
WELL00016201804261451 08	0,6638	0,8024	0,7265
WELL00016201805172223 22	0,7661	0,8569	0,8090
WELL00017201403171517 43	0,5722	0,7340	0,6431
WELL00017201403180231 41	0,5283	0,6834	0,5959
WELL00017201403181602 20	0,4212	0,6720	0,5178
WELL00017201403190404 53	0,5873	0,6384	0,6118
WELL00017201403191414 50	0,5248	0,6268	0,5713
WELL00002201402121703 33	0,8057	0,7629	0,7837
WELL00002201403011517 00	0,7665	0,5374	0,6318
WELL00002201403251703 04	0,7788	0,7268	0,7519
WELL00004201710311815 09	0,8657	0,7618	0,8104
WELL00004201710311930 25	0,7749	0,5016	0,6090
WELL00004201710312000 59	0,7903	0,5218	0,6286
WELL00001201702262203 09	0,4656	0,6204	0,5320
WELL00006201806181107 21	0,4042	0,4221	0,4130
WELL00006201806201813 48	0,7041	0,6315	0,6658
WELL00021201705090135 17	0,6936	0,5948	0,6404
WELL00020201204101923 26	0,4658	0,7034	0,5605
WELL00019201703011823 17	0,6574	0,6193	0,6378

Fonte: própio autor.

Tabela 14 – Resultados dados 3W para o modelo Floresta de Isolamento, *recall*, *precission* e *F1 score* com 500 atrasos.

Poço	*recall*	*precision*	*F1 score*
WELL00001201401242131 36	1,000	0,9775	0,9886
WELL00002201401262000 50	1,000	0,9965	0,9982
WELL00006201708010636 14	0,9487	0,6124	0,7443
WELL00006201708021230 00	0,4366	0,7154	0,5423
WELL00006201806180602 45	1,000	0,9903	0,9951
WELL00002201311040141 01	0,9867	0,9837	0,9852
WELL00009201703131608 04	1,000	0,9924	0,9962
WELL00010201712182001 31	1,000	0,9987	0,9993
WELL00015201706201603 49	0,9352	1,000	0,9665
WELL00015201710131400 47	0,7145	0,9563	0,8179
WELL00016201804050203 45	0,9726	0,7355	0,8376
WELL00016201804261420 05	1,000	0,6662	0,7997
WELL00016201804261451 08	0,7152	0,8530	0,7780
WELL00016201805172223 22	0,9634	1,000	0,9814
WELL00017201403171517 43	0,6012	0,9725	0,7430
WELL00017201403180231 41	0,5412	0,8418	0,6588
WELL00017201403181602 20	0,5674	0,8338	0,6753
WELL00017201403190404 53	0,7371	0,9521	0,8309
WELL00017201403191414 50	0,7415	0,9934	0,8492
WELL00002201402121703 33	0,9724	0,8398	0,9012
WELL00002201403011517 00	1,000	0,6634	0,7976
WELL00002201403251703 04	0,7788	0,9837	0,8693
WELL00004201710311815 09	0,9924	0,8509	0,9162
WELL00004201710311930 25	1,000	0,6845	0,8127
WELL00004201710312000 59	1,000	0,6257	0,7698
WELL00001201702262203 09	0,6620	0,9824	0,7910
WELL00006201806181107 21	0,4627	0,5682	0,5101
WELL00006201806201813 48	0,9874	0,8636	0,9214
WELL00021201705090135 17	0,9934	0,9233	0,9571
WELL00020201204101923 26	0,5904	0,8427	0,6943
WELL00019201703011823 17	0,6807	0,8169	0,7425

Fonte: própio autor.