



**PROGRAMA FRANCISCO EDUARDO MOURÃO SABOYA DE  
PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA**

**ESCOLA DE ENGENHARIA**

**UNIVERSIDADE FEDERAL FLUMINENSE**

**Dissertação de Mestrado**

**DETECÇÃO DE PROBLEMAS DE  
GARANTIA DE ESCOAMENTO A PARTIR  
DA UTILIZAÇÃO DE FERRAMENTAS DE  
APRENDIZADO DE MÁQUINA**

**MAYARA DE JESUS ROCHA SANTOS**

**OUTUBRO DE 2020**

MAYARA DE JESUS ROCHA SANTOS

DETECÇÃO DE PROBLEMAS DE GARANTIA DE  
ESCOAMENTO A PARTIR DA UTILIZAÇÃO DE  
FERRAMENTAS DE APRENDIZADO DE MÁQUINA

Dissertação de Mestrado apresentada  
ao Programa Francisco Eduardo Mourão  
Saboya de Pós-Graduação em Engenharia  
Mecânica da UFF como parte dos requisitos  
para a obtenção do título de Mestre em  
Ciências em Engenharia Mecânica

Ficha catalográfica automática - SDC/BEE  
Gerada com informações fornecidas pelo autor

S237d Santos, Mayara de Jesus Rocha  
DETECÇÃO DE PROBLEMAS DE GARANTIA DE ESCOAMENTO A PARTIR DA  
UTILIZAÇÃO DE FERRAMENTAS DE APRENDIZADO DE MÁQUINA / Mayara  
de Jesus Rocha Santos ; Fabiana Rodrigues Leta, orientador.  
Niterói, 2020.  
117 f.  
  
Dissertação (mestrado)-Universidade Federal Fluminense,  
Niterói, 2020.  
  
DOI: <http://dx.doi.org/10.22409/PGMEC.2020.m.14116171778>  
  
1. Garantia de escoamento. 2. Algoritmos de classificação.  
3. Poços offshore. 4. Produção intelectual. I. Leta,  
Fabiana Rodrigues, orientador. II. Universidade Federal  
Fluminense. Escola de Engenharia. III. Título.  
  
CDD -

Orientadora: Fabiana Rodrigues Leta (PGMEC/UFF)

UNIVERSIDADE FEDERAL FLUMINENSE  
NITERÓI, 13 DE OUTUBRO DE 2020

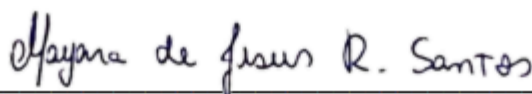
# DETECÇÃO DE PROBLEMAS DE GARANTIA DE ESCOAMENTO A PARTIR DA UTILIZAÇÃO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

Esta Dissertação é parte dos pré-requisitos para a obtenção do título de

## **MESTRE EM ENGENHARIA MECÂNICA**

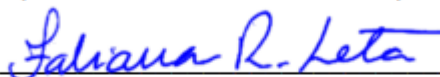
Área de concentração: Termociências

Aprovada em sua forma final pela Banca Examinadora formada pelos professores:



---

MAYARA DE JESUS ROCHA SANTOS  
(Candidata ao título de Mestre)



---

PROF<sup>a</sup>. FABIANA RODRIGUES LETA (Orientadora)  
(PGMEC/UFF)



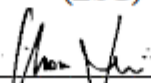
---

PROF<sup>a</sup>. MARIA DA PENHA CINDRA FONSECA (Membro Interno)  
(PGMEC/UFF)



---

PROF. LUIZ OCTÁVIO GAVIÃO (Membro Externo)  
(ESG)



---

PROF. GILSON BRITO ALVES LIMA (Membro Extra)  
(UFF)

Aos Meus Pais e Meus Dindos.

# Agradecimentos

A minha mãe, Maria, pelo suporte e apoio durante essa jornada e a minha dinda, Marluci, por me proporcionar a honra de ser “filha da dinda”.

Ao meu pai, Edvaldo, e ao meu dindo, Raimundo, que apesar de não estarem mais neste plano continuam perto, em meu coração, e me acompanham nesse caminho.

As minhas orientadora e “madrinha”, Fabiana e Maria, pelo apoio durante a realização desse projeto. Foi muito gratificante poder contar com duas mulheres inteligentes que, sem dúvidas, são exemplos a serem admirados.

Aos meus professores e colegas de projeto pelo suporte, pelas dúvidas tiradas, pelas dicas, pelo aprendizado e pela paciência infinita. Agradeço, em especial, ao Geraldo que sempre esteve presente na minha vida acadêmica e é um querido mentor e amigo.

Ao meu companheiro, Marcos, que dividiu comigo essa experiência e que traz amor e alegria para os meus dias.

Ao meu amigo Pedro que continua sendo minha referência de pessoa mais inteligente do mundo e agora é também um dos melhores professores do universo.

Aos meus irmãos e primos, por fazerem parte da minha vida mesmo à distância.

Aos colaboradores do PGMEC e a CAPES.

Por fim, gostaria de agradecer todo o suporte fornecido pela Petrobras, através do projeto de pesquisa junto à ANP e UFF nº 21354-6, ao longo do desenvolvimento deste trabalho.

# Resumo

Na produção de poços *offshore* podem ocorrer os denominados problemas de garantia de escoamento. Esses problemas podem afetar tubulações e equipamentos e, em casos mais severos, interromper a produção de óleo e gás. Alguns exemplos são formação de hidratos em linhas de produção, incrustação em válvulas, *slugging* e instabilidade de fluxo. No intuito de detectar a ocorrência dessas anomalias durante a produção de petróleo e estabelecer meios que permitam diferenciar anomalias com impactos similares na produção, este trabalho propõe combinar o estudo do comportamento dos sensores de pressão e temperatura localizados no sistema de produção com a aplicação de ferramentas de aprendizado de máquina como; *adaboost*, árvore de decisão, modelos de misturas gaussianas, *random forest* e *naive Bayes*. Esses algoritmos de classificação foram aplicados para identificar diferentes classes presentes em um banco de dados público contendo dados reais de produção de diversos poços de petróleo. Por meio dos experimentos realizados nesta pesquisa, também verificou-se a possibilidade de utilizar modelos classificadores de anomalias para mais poços além do usado para treinar o modelo. Todavia, apenas modelos de classificação de formação de hidrato apresentaram resultados positivos quando usados em poços diferentes. Os algoritmos de aprendizado de máquina mostraram bons resultados na classificação dos poços em regimes normais, transientes e anormais, onde a falha já está instalada, obtendo *F1-score* entre 97 e 100%. Resultados positivos também foram obtidos na classificação de dados de *slugging* e instabilidade de fluxo, os modelos foram capazes de diferenciar esses dois problemas com *F1-score* 99 a 100%.

**Palavras-chave:** garantia de escoamento, algoritmos de classificação, poços offshore

# Abstract

In offshore fields a main concern is flow assurance issues. Flow assurance can impact equipment, production flow and, in severe cases, shut down the production of oil and gas. Hydrate formation, scaling, slugging and flow instability are some of flow assurance problems that occurs in petroleum production. In order to identify the occurrence of such problems, classify well production status and distinguish similar anomalies, the present work proposes a detailed analysis of pressure and temperature sensors combined with machine learning techniques like adaboost, decision tree, random forest, naïve Bayes and gaussian mixture models. Thus, these classifiers were applied for detecting different classes such as “normal”, “transient” and “abnormal” in a public dataset with real data about undesirable anomalies during oil production. Also, this work has runned experiments designed to check the possibility of using models based on one well to classify different others wells, and only for hydrate formation positive results were achieved in this matter. Nevertheless, the machine learning classifiers showed high F1-scores for well status classification scoring between 97-100%. Likewise, for slugging and flow instability classification the algorithms presented F1-score up to 99%, reaching 100% in most cases.

**Key words:** flow assurance, classifying algorithm, offshore wells



# Sumário

Capítulo 1 .....	17
1.1. Contextualização .....	19
1.2. Problema de pesquisa .....	21
1.3. Questões de pesquisa.....	21
1.4. Objetivos.....	22
1.5. Justificativa .....	22
1.6. Delimitação da pesquisa .....	23
1.7. Contribuição científica .....	23
1.8. Da organização da dissertação .....	24
Capítulo 2.....	25
2.1. Produção <i>Offshore</i> .....	25
2.1.1. Sensores .....	27
2.2. Garantia de escoamento .....	28
2.2.1. Hidratos de gás natural .....	29
2.2.2. Incrustação .....	34
2.2.3. Instabilidade de fluxo .....	37
2.2.4. <i>Slugging</i> .....	39
Capítulo 3.....	43
3.1. Mineração de dados e aprendizado de máquina .....	43
3.2. Aprendizado supervisionado e não supervisionado.....	44
3.3. Métodos de partição do conjunto de dados .....	45
3.4. Algoritmos de classificação.....	47
3.4.1. Adaboost.....	48
3.4.2. <i>Naive Bayes</i> .....	49
3.4.3. Modelos de Misturas Gaussianas (GMM).....	50
3.4.4. Árvore de decisão.....	51

3.4.5. <i>Random forest</i> .....	54
3.5. Métricas de avaliação .....	55
Capítulo 4 .....	58
4.1. Metodologia experimental bloco 1 .....	63
4.2. Metodologia experimental do bloco 2 .....	65
Capítulo 5 .....	66
5.1. Bloco de experimentos formação de hidrato e incrustação .....	66
5.1.1. Análise dos poços .....	66
5.1.2. Experimento 1 .....	75
5.1.3. Experimento 2 .....	80
5.1.4. Experimento 3 .....	81
5.1.5. Experimento 4 .....	85
5.1.6. Discussão dos resultados dos experimentos do bloco 1 .....	86
5.2. Bloco de experimentos <i>slugging</i> e instabilidade de fluxo .....	87
5.2.1. Experimento 1: análise temporal .....	88
5.2.2. Experimento 2 .....	100
5.2.3. Experimento 3 .....	102
5.2.4. Discussão dos resultados dos experimentos do bloco 2 .....	103
Capítulo 6 .....	106
6.1. Conclusões .....	109
6.2. Desafios futuros .....	110
Referências .....	112

## Lista de figuras

Figura 1- Esquema de uma Unidade Estacionária de Produção .....	25
Figura 2- Arquitetura submarina de um sistema de produção <i>offshore</i> .....	26
Figura 3– <i>Plug</i> de hidrato dentro de uma linha de produção .....	30
Figura 4- Processo de formação de hidratos .....	31
Figura 5- Pontos de maior risco de formação de hidratos em sistemas offshore	33
Figura 6- Exemplo de incrustação em uma linha de produção em terra .....	35
Figura 7 -Padrões de escoamento em tubulações verticais, de fluxo ascendente, e horizontais .....	39
Figura 8- Esquema de formação de slug.....	41
Figura 9-Processo de descobrimento de conhecimento (Knowledge discovery)	44
Figura 10- Esquema de validação cruzada usando k- folds.....	46
Figura 11- Esquema de partição de conjunto de treino e teste.....	47
Figura 12- Esquema lógico do algoritmo Adaboost.....	49
Figura 13- Exemplo de árvore de decisão genérica.....	53
Figura 14- Algoritmo Random Forest .....	55
Figura 15-Estatística descritiva do poço 20 .....	67
Figura 16- Estatística descritiva do poço 21 .....	67
Figura 17- Comparação temporal das variáveis de produção do poço 20. A linha laranja indica o início do transiente e a linha vermelha indica a instalação da falha. ....	68
Figura 18- Comparação temporal das variáveis de produção do poço 21. A linha laranja indica o início do transiente e a linha vermelha indica a instalação da falha. ....	70
Figura 19- Estatística descritiva do poço 01 para problema de incrustação na válvula de produção .....	71
Figura 20- Estatística descritiva do poço 06 .....	71
Figura 21-Comparação temporal entre as variáveis do poço 01. A linha laranja indica o início do transiente e a linha vermelha indica a instalação da falha.....	73
Figura 22- Comparação temporal entre as variáveis do poço 06, onde a linha laranja indica o início do transiente. ....	74
Figura 23-Matriz de confusão do modelo gerado pelo GMM .....	76

Figura 24- Exemplo de gráfico gerado pelo algoritmo naive Bayes .....	77
Figura 25- Matriz de confusão do naive Bayes.....	77
Figura 26- Matriz de confusão do modelo gerado pelo Adaboost. ....	78
Figura 27-Matriz de confusão da árvore de decisão para problemas de formação de hidrato.....	78
Figura 28-Árvore de decisão para formação de hidratos no poço 20.....	79
Figura 29- Importância de cada variável para classificação da variável alvo no conjunto de treino do poço 20 .....	80
Figura 30- Resultado da aplicação dos modelos treinados com dados do poço 20 para classificação do poço 21.....	80
Figura 31- Matriz de confusão relacionada ao GMM para o poço 01.....	82
Figura 32- Matriz de confusão referente ao algoritmo naive Bayes usado nos dados do poço 01 .....	82
Figura 33- Matriz de confusão referente ao algoritmo Adaboost usado nos dados do poço 01 .....	83
Figura 34- Árvore de decisão do poço 01 .....	84
Figura 35- Matriz de confusão do modelo gerado pela árvore de decisão .....	84
Figura 36- Importância de cada variável para classificação da variável alvo no conjunto de treino do poço 01 .....	85
Figura 37- Resultado da aplicação dos modelos treinados com dados do poço 1 para classificação do poço 6.....	86
Figura 38- Sensor de pressão na árvore de natal (P.TPT) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo.....	90
Figura 39- Sensor de temperatura na árvore de natal (T.TPT) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo.....	91
Figura 40- Sensor de pressão à montante da válvula de produção (P.MON.CKP) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo.....	92
Figura 41- Sensor de temperatura à jusante da válvula de produção (T.JUS.CKP) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo.....	93
Figura 42- Sensor de pressão à jusante da válvula de injeção de gas lift (P.JUS.CKGL) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo .....	95
Figura 43- Sensores de pressão durante a produção normal do poço.....	96
Figura 44- - Comportamento dos sensores de pressão durante slugging.....	97

Figura 45- Comportamento dos sensores de pressão durante a instabilidade de fluxo .....	98
Figura 46- Comportamento dos sensores de temperatura durante os cenários “normal”, “slugging” e “instabilidade de fluxo”. .....	99
Figura 47- Resultado da aplicação dos algoritmos de classificação nos dados de slugging e instabilidade de fluxo concatenados .....	100
Figura 48- Árvore de decisão para classificação de dados de slugging e instabilidade.....	101
Figura 49- Importância das variáveis para a construção do modelo de classificação da random forest .....	102
Figura 50- Resultado obtidos pelos modelos criados a partir do poço 01 para classificação dos dados de slugging e instabilidade de fluxo no poço 14.....	103

## **Lista de tabelas**

Tabela 1- Matriz de confusão.....	56
Tabela 2- Identificação de sensores e variáveis relacionadas .....	59
Tabela 3– Experimentos da dissertação.....	61
Tabela 4- Pacotes em R necessários para a realização dos experimentos .....	64
Tabela 5- Resumo da variação de amplitude de cada variável em cada poço analisado, usando a diferença entre o maior e o menor valor disponível. ....	75
Tabela 6 – Resumo dos resultados obtidos com o classificador naive Bayes para cada sensor do poço 20 .....	77
Tabela 7- Resumo dos resultados obtidos com o classificador naive Bayes para cada sensor do poço 01 .....	82
Tabela 8 - Média das variáveis observadas em cada situação estudada para o poço 01 .....	88

## Lista de abreviaturas

**ANM-** Árvore de Natal Molhada

**ANP-** Agência Nacional de Petróleo, Gás Natural e Biocombustíveis

**CART-** *Classification and Regression Tree*

**GMM-** *Gaussian Mixture Models* (Modelos de Misturas Gaussianas)

**IA-** Inteligência Artificial

**IEA-** *International Energy Agency*

**KDE-** *Kernel Density Estimation*

**KNN-** *K-Nearest Neighbors*

**PCA-** *Principal Components Analysis* (Análise de Componentes Principais)

**SVM-** *Support Vector Machine*

**UEP-** Unidade Estacionária de Produção

## Glossário

**Workover:** é desenvolver em um poço de óleo e gás um processo de manutenção ou ação de remediação de grande proporção. Implica na remoção e recolocação da coluna de produção, onde há interrupção da produção durante o procedimento. Além disso, demanda que sondas especializadas em intervenção de poços sejam utilizadas (OILFIELD GLOSSARY, 2020).

**Matriz de confusão:** é uma matriz que mostra desempenho do modelo de classificação. Nela são apresentados os dados positivos, falsos positivos, negativos e falsos negativos. É com base nessa matriz que são calculadas métricas como *precision*, *recall* e *F1-score*, por exemplo (MISHRA, 2018).

**Precision:** é a métrica que computa a quantidade de instâncias classificadas como verdadeiros positivos dividida pelo total de instâncias classificadas como positivas, verdadeiros positivos e falsos positivos, (MISHRA, 2018).

**Recall:** é a métrica que computa a quantidade de instâncias classificadas como verdadeiros positivos dividida pelo total de instâncias que deveriam ser classificadas como positivas, verdadeiros positivos e falsos negativos, (MISHRA, 2018).

**Acurácia:** é uma métrica que contabiliza a quantidade de predições corretas pelo total de amostras preditas (MISHRA, 2018).

**F1-score:** é a métrica que representa a média harmônica entre *recall* e *precision* e proporciona melhor entendimento sobre quão preciso e robusto é o modelo analisado. Sendo assim, quanto maior seu valor melhor é o desempenho do modelo (MISHRA, 2018).

**Manifold:** é um equipamento que pode ser constituído por tubos ou válvulas. Pode apresentar diversas funções, como distribuir e monitorar o fluxo de fluido produzido e ligar a produção de poços diferentes em uma linha principal conectada a unidade de produção (OILFIELD GLOSSARY, 2020).

**Riser:** é o tubo que liga a unidade de produção *offshore* ao sistema de produção submarino. Na produção, é responsável por conectar as linhas de produção à unidade de produção. Na perfuração, por outro lado, liga a coluna de perfuração à plataforma de perfuração (PANORAMA OFFSHORE, 2019)



# Capítulo 1

## Introdução

O monitoramento da produção de óleo e gás é de suma importância para assegurar a manutenção do processo produtivo e a redução dos custos em operações de intervenção em poços, chamadas de *workovers*. Na produção *offshore*, o monitoramento se torna uma atividade ainda mais importante pois os riscos e custos envolvidos são em geral muito maiores, visto que a produção de campos marítimos conta com equipamentos e sistemas mais complexos e, se tratando de volume produzido, é muito maior quando comparados aos de campos terrestres (VARGAS *et al*, 2017). No caso do Pré-sal, por exemplo, três campos correspondem por mais da metade da produção nacional (ANP, 2019).

Novas tecnologias envolvendo IA (inteligência artificial) e Aprendizado de Máquina vêm sendo aplicadas em diversas indústrias, e a indústria petrolífera não é exceção. Durante operações de perfuração e de produção de poços, milhares de informações são geradas a cada segundo. Por essa grande disponibilidade de dados, a indústria de petróleo é um nicho potencial para aplicações de algoritmos inteligentes, capazes de extrair informações importantes da base de dados (SANTOS *et al*, 2019).

É cada vez mais recorrente o uso de algoritmos de inteligência para auxiliar o processo de detecção de sólidos e furo de coluna na fase de produção, bem como detectar eventos de prisão de coluna na fase de perfuração e ajudar no diagnóstico de outros problemas. Exemplos de aplicações em perfuração de poços podem ser encontrados em Alouhali *et al* (2018), Pinto (2018), Marques (2019), Pereira (2019) e Khodnenko *et al* (2019). Akinsete e Oshingbesan (2019), Qin *et al* (2019), Santos *et al* (2018) apresentam aplicações recentes de ferramentas de inteligência artificial na fase de produção. A quantidade de dados gerados durante toda vida produtiva do poço é enorme e com o auxílio da inteligência artificial é possível extrair ainda mais informações desses dados. A crescente aplicação de inteligência artificial no setor, impulsionada pela “Indústria 4.0” está promovendo novas abordagens para problemas antigos. No caso de monitoramento de hidratos, que é o principal problema de garantia de escoamento podem ser aplicados, por exemplo, algoritmos como redes neurais para auxiliar o monitoramento da produção de óleo e gás dentro da zona de estabilidade de hidratos (GLÉNAT, 2015).

A seguir, é apresentada uma revisão da literatura a cerca da aplicação de técnicas de IA e aprendizado de máquina no setor de óleo e gás. Serão apresentadas resumidamente as técnicas que vêm sendo aplicadas e os resultados obtidos pelos autores. Posteriormente serão apresentados trabalhos usando metodologias convencionais, como o uso de simulações e correlações, em problemas de garantia de escoamento.

Akinsete & Oshingbesan (2019) utilizam modelos inteligentes para detectar vazamentos em linhas de gás natural. Os autores usaram *gradient boosting*, *decision trees*, *random forest*, *support vector machine* e redes neurais artificiais como modelos inteligentes. Dentre esses, *random forest* e *decision tree* apresentaram os melhores resultados, conseguindo detectar vazamentos de até 0,1% da vazão nominal.

Outros artigos na literatura apresentam resultados positivos para a aplicação de *random forest* na detecção de anomalias durante a produção de petróleo. Santos *et al* (2018) fazem uso do algoritmo para detectar hidratos em linhas de produção e de injeção e obtém um *F1-score* de 0,76. Nesse mesmo trabalho, os autores, propõem também o uso de modelos como Análise de Componentes Principais (PCA) e *Gaussian Mixture Models* (GMM) para auxiliar a classificação e o tratamento de dados.

Outra aplicação de *random forest* na detecção de hidratos na linha de produção é apresentada por Marins (2018), em sua dissertação. O autor propõe um sistema de classificação dividido em três blocos: o primeiro extrai informações estatísticas dos dados de entrada por meio de uma janela deslizante; o segundo mapeia a saída do bloco anterior utilizando as transformações *z-score* e PCA; por último, o bloco de classificação usa o *random forest* como classificador.

Por outro lado, há também aplicações de *random forest* para detectar a prisão mecânica da coluna perfuração, como é apresentado por Pereira (2019). O autor, por intermédio de um conjunto de variáveis, classifica se a coluna está efetivamente perfurando ou se está, de fato, presa. Os resultados obtidos por meio da utilização do algoritmo de *random forest* foram positivos. A técnica obteve 94% de acurácia na classe “preso” e 71% para classe “não preso”.

Khodnenko *et al* (2019) desenvolveram um trabalho sobre a aplicação de técnicas de aprendizado de máquina na detecção de perda de circulação durante a perfuração de poços. Os autores aplicaram regressão logística, *support vector machine* (SVM), *k-nearest neighbors* (KNN), *Naive Bayes* e *random forest*. Os dados de entrada para os modelos foram a taxa de fluido injetado durante a perfuração e o *water cut* (referente à razão de água produzida e o volume total de líquidos produzidos). Dentre os métodos

aplicados o de melhor desempenho foi a regressão logística, apresentando 79% de acurácia.

Alouhali *et al* (2018) propõem uma forma automatizada de detectar *kick*, que ocorre quando fluidos da formação adentram o poço durante a perfuração (OILFIELD GLOSSARY, 2020), usando mineração de dados. Cinco modelos foram desenvolvidos neste trabalho: árvore de decisão, KNN, *Sequential Minimal Optimization* (SMO), redes neurais artificiais e redes bayesianas. Os modelos foram treinados com dados reais de *kicks* e os modelos que apresentaram os melhores resultados foram a árvore de decisão e o KNN.

Qin *et al* (2019) avaliam o risco de formação de *plugs* de hidratos durante a produção de petróleo usando dados de campo e dados de laboratório. Para auxiliar o monitoramento da produção dentro do envelope de formação de hidratos, os autores utilizam SVM e redes neurais como modelos de análise do banco de dados para determinar a fração de hidrato na linha. Com dados de laboratório e dados reais, os modelos apresentaram bom desempenho e foram aplicados em campos reais.

Na literatura há diversos registros de aplicações de técnicas de IA para resolver problemas recorrentes da indústria de óleo e gás. No que tange a problemas de garantia de escoamento, havia o uso recorrente de correlações e simulações para determinar as condições físicas e termodinâmicas em que tais problemas ocorrem e, assim, os profissionais pudessem evitá-los. Tais metodologias continuam sendo aplicadas na modelagem, detecção e predição de problemas de incrustação, deposição de parafina e outros sólidos (COTO *et al*, 2010; BALAKIN *et al*, 2016; DUAN *et al*, 2017; SHI *et al*, 2018; CHAUDHARI *et al*, 2018; ALNAIMAT & ZIAUDDIN, 2020). No entanto, com o avanço das ferramentas de monitoramento, a forma de lidar com um dos principais problemas de garantia de escoamento, a formação de hidratos, vem mudando. Antes, as operadoras trabalhavam fora da faixa de formação do hidrato a todo custo. Agora, visto que pode ser mais rentável, empresas trabalham dentro da faixa de formação de hidratos e, ao invés de evitá-los, elas gerenciam a sua formação, de forma a garantir que o hidrato nunca se aglomere e feche completamente a linha de produção (GLÉNAT *et al*, 2015).

### **1.1.Contextualização**

Durante a produção de hidrocarbonetos podem ocorrer problemas relacionados ao escoamento dos fluidos. Esses são chamados problemas de garantia de escoamento e estão intrinsecamente relacionados à produção *offshore*. Na produção de poços marítimos,

existem diversos fatores que tornam o escoamento da produção mais complexo. A conjuntura de quilômetros de tubulações dispostas sobre o irregular solo marinho, temperaturas baixas e a presença de diversos equipamentos fazem com que o entendimento do comportamento do fluxo seja primordial para garantir que a produção de fluidos ocorrerá de forma eficiente. Um dos principais postos-chave da garantia de escoamento na indústria de petróleo é a deposição de sólidos nas linhas de produção e transporte, tais como asfaltenos, parafinas, hidratos de gás e naftenos (GUDMUNDSSON, 2018). Dentre estes, os hidratos de gás são considerados os mais importantes. Hidratos podem se formar rapidamente e sem apresentar indícios, se comparado à deposição de asfaltenos, parafinas e incrustação. Além disso, podem levar dias ou meses para serem solucionados (SLOAN *et al*, 2009).

A produção e transporte de petróleo e seus derivados dependem também do conhecimento acerca dos padrões de escoamento multifásico. O regime de escoamento está relacionado à geometria da linha de produção, a propriedades dos fluidos envolvidos e às condições de fluxo. O mais crítico dentre os regimes de escoamento é o *slugging*, caracterizado por golfadas periódicas de gás que impactam significativamente a queda de pressão ao longo da tubulação. Com isso, a presença de *slug* diminui a eficiência do transporte de fluidos e também pode acarretar danos físicos aos equipamentos (CAMPBELL & LIU, 2016).

O desenvolvimento de campos *offshore* conta com investimentos em tecnologias que possibilitam enviar informações em tempo real de poços localizados em regiões remotas para centros de monitoramento em terra de maneira eficiente (LIMA, 2018). Sendo assim, informações sobre falhas e anomalias na produção de óleo podem ser acessadas em terminais de monitoramento *onshore*, fazendo com que o processo decisório deixe de pertencer apenas aos profissionais de campo.

O cenário atual de algumas operadoras no Brasil é de identificar a falha e confirmar o diagnóstico depois. Em determinados casos, um profissional é responsável por monitorar diversos parâmetros ao mesmo tempo, por meio de um sistema integrado de monitoramento, contendo informações de todos os sensores, de todos os poços de uma unidade estacionária de produção. Quando o sistema alarma alguma anormalidade o profissional deve analisar a conjuntura de falha e realizar as medidas cabíveis. Posteriormente, o primeiro diagnóstico é confirmado ou não. Nesse caso, um algoritmo inteligente poderia ser treinado e testado para detectar os primeiros estágios do problema antes que ele se desenvolva completamente e, também, estabelecer meios para um

diagnóstico mais preciso sobre o tipo de problema acontecendo. Dessa forma, o processo de monitoramento e remediação de falhas poderia ser otimizado, auxiliando o profissional responsável pelo monitoramento dos poços na tomada de decisões.

Nesse sentido, o presente estudo visa colaborar na detecção de hidratos, incrustações, instabilidades de fluxo e *slugging* através da caracterização e distinção de suas respectivas respostas de sensores de pressão e temperatura comumente encontrados num sistema de produção.

### **1.2.Problema de pesquisa**

O problema abordado nesta pesquisa é como diferenciar problemas de garantia de escoamento como *slugging*, instabilidade de fluxo, incrustação e hidrato tendo como ponto de partida dados de um conjunto de sensores localizados no leito marinho e no fundo do poço, bem como na unidade estacionária de produção. A partir do estudo do comportamento dos dados referentes a cada problema, espera-se que seja possível identificar suas respectivas especificidades, independentemente do conhecimento prévio acerca das condições físicas que envolvem um sistema de produção *offshore*. Dessa forma, será possível realizar o diagnóstico do problema antes que o mesmo se instale definitivamente, sendo muito importante para redução de custos com operações de intervenção.

### **1.3.Questões de pesquisa**

A presente pesquisa visa estabelecer diferenças entre os problemas de garantia de escoamento por meio da análise de sensores de produção. Para caracterizar essas anomalias durante a produção de óleo e gás, é importante responder as seguintes questões:

- Quais sensores são determinantes para caracterizar problemas como formação de hidrato e incrustação? Visto que ambos causam restrições ao diâmetro disponível ao escoamento.
- Quais sensores são determinantes para diferenciar ou classificar a ocorrência de *slugging* e de instabilidade de fluxo?
- Como é o desempenho de algoritmos de aprendizado de máquina na classificação de dados reais com diferentes anomalias?
- Um mesmo modelo de classificação pode ser usado em mais de um poço?

## 1.4. Objetivos

O objetivo principal desta dissertação é distinguir problemas de garantia de escoamento que ocorrem durante a produção de petróleo e gás natural, como: *slugging*, instabilidade de fluxo, incrustação e hidrato, sob a ótica da ciência de dados, com a utilização de técnicas de inteligência artificial como o aprendizado de máquina.

A partir de um banco de dados reais disponível na literatura (VARGAS *et al*, 2019), busca-se entender como a natureza e a variabilidade dos dados fornecem informações suficientes para que seja feita a distinção entre esses tipos de problemas. Portanto, deseja-se entender como profissionais que analisam esses dados em tempo real distinguem qual problema está acontecendo naquele instante. Sendo assim, com a análise dos dados reais da literatura por meio de ferramentas de inteligência artificial será possível estabelecer os critérios ou atributos que diferenciam cada problema.

Como objetivo secundário almeja-se determinar a viabilidade da aplicação de algoritmos de classificação como Gaussian Mixture Model (GMM), Árvore de decisão Random Forest, Naive Bayes e Adaboost para classificar corretamente os problemas de garantia de escoamento abordados. Uma vez que *slugging* e instabilidade de fluxo possuem características físicas semelhantes, assim como incrustação e hidrato, é esperado que a análise dos dados permita entender como as respostas dos sensores são diferentes para cada problemas apesar da similaridade quanto às suas características físicas.

## 1.5. Justificativa

A justificativa para esta pesquisa baseia-se na necessidade de continuar melhorando os métodos e ferramentas para aumentar a eficiência do processo de detecção de falhas e anomalias. Na indústria de óleo e gás, cada operação de intervenção, chamada de *workover*, pode custar milhões de dólares. Atualmente, em algumas empresas, existe somente um operador responsável pelo monitoramento de diversos painéis contendo inúmeras informações acerca dos poços produtores. É de se esperar que mudanças sutis em determinados parâmetros não sejam detectadas imediatamente, porém com a evolução dos problemas de garantia de escoamento, as variáveis começam a apresentar variações significativas que levam o operador a tomar as devidas atitudes. No entanto, apesar dos profissionais tomarem as medidas necessárias, o diagnóstico definitivo é realizado

posteriormente ao evento. Com o auxílio dos algoritmos de inteligência artificial é possível detectar padrões e variações de maneira mais rápida e precisa que a mente humana. Nesse sentido, esta pesquisa objetiva empregar tais ferramentas para auxiliar o processo de diagnóstico de problemas como instabilidade de fluxo, *slugging*, incrustação e hidrato e, por meio de algoritmos classificadores, realizar o diagnóstico antes que problema se instale de maneira irreversível. Com isso, além de auxiliar os profissionais responsáveis pelo monitoramento dos poços, a utilização desses algoritmos pode reduzir custos com *workovers*, e diminuir o tempo necessário para identificação de problemas.

### **1.6.Delimitação da pesquisa**

A obtenção de dados reais de produção é difícil, pois essas informações são consideradas sensíveis para as empresas operadoras. Por essa razão, optou-se por utilizar o banco de dados publicado por Vargas *et al* (2019), com dados de diversos poços e anomalias. Também, para aplicação de técnicas de aprendizado supervisionado, é necessário ter a “classe” do atributo alvo, ou seja, é preciso que os dados já estejam com as marcações referentes aos tipos de falhas ocorridas. Com dados não classificados, ou obtidos em tempo real, não seria possível realizar as análises presentes nessa pesquisa.

Os problemas de garantia de escoamento abordados também foram delimitados pela sua disponibilidade no banco de dados. Não havia, por exemplo, dados referentes à deposição de parafina ou asfalteno. Assim como não havia informações sobre outros regimes de escoamento. É possível que os poços disponíveis no banco de dados não apresentassem tais ocorrências, e, por isso, não havia documentação.

### **1.7.Contribuição científica**

A implementação de algoritmos de classificação em dados reais de poços contribui para uma melhor documentação de falhas e anomalias durante o processo de produção de petróleo. Com isso, é possível alocar melhor os recursos e indentificar as falhas que são mais recorrentes e, assim, traçar estratégias mais eficientes de combate às anomalias. Especificamente, essa dissertação permite identificar variáveis reais de produção que têm potencial para caracterizar problemas de garantia de escoamento como incrustação, formação de hidrato, *slugging* e instabilidade de fluxo. A identificação e compreensão do comportamento dos sensores de produção durante a ocorrência dessas

anomalias proporcionam conhecimento para pessoas que não possuem a experiência de profissionais que trabalham diariamente com monitoramento de poço.

### **1.8. Organização da dissertação**

Esta dissertação contém seis capítulos. O primeiro contempla a introdução, a contextualização do problema abordado, o problema de pesquisa, os objetivos e a justificativa de pesquisa. O capítulo dois aborda a revisão teórica de conceitos relacionados à engenharia de petróleo e mecânica dos fluidos, necessários para plena compreensão do trabalho. O terceiro capítulo apresenta a revisão teórica de conceitos da área de inteligência artificial, aprendizado de máquina e seus algoritmos de classificação. No capítulo quatro é descrita a metodologia proposta e os experimentos a serem realizados. Em seguida, no capítulo cinco, os resultados e discussões acerca dos experimentos realizados são percorridos. Por último, são apresentadas as conclusões da pesquisa, bem como sugestões para trabalhos futuros.



# Capítulo 2

## Revisão teórica – conceitos da área de Engenharia de Petróleo

Neste capítulo será realizada uma revisão teórica sobre temas e conceitos da área de Engenharia de Petróleo que são importantes para entender o objetivo, a área de aplicação e o contexto no qual esta pesquisa está inserida.

### 2.1. Produção *Offshore*

Com a descoberta de reservatórios de petróleo e gás com grande potencial de retorno de investimentos, sob o leito marinho, a Indústria de Petróleo passou a contar com uma importante área de exploração e produção: o mar. Os campos *offshore*, no Brasil, são responsáveis por mais da metade da produção nacional (ANP, 2020). No cenário internacional, mais de 25% da produção de óleo e gás é *offshore* (IEA, 2018).

A produção de um poço *offshore*, assim como um *onshore*, se inicia após a finalização das etapas de perfuração e completação. É na completação que é instalado um dos principais equipamentos de controle e segurança do poço, a árvore de natal. A depender do posicionamento da árvore de natal, a completação é classificada como seca – árvore de natal posicionada na superfície ou molhada-árvore de natal localizada sobre o solo marinho (LAIK, 2018). Neste estudo, a árvore de natal molhada (ANM) está localizada no leito marinho, como mostra a Figura 1.

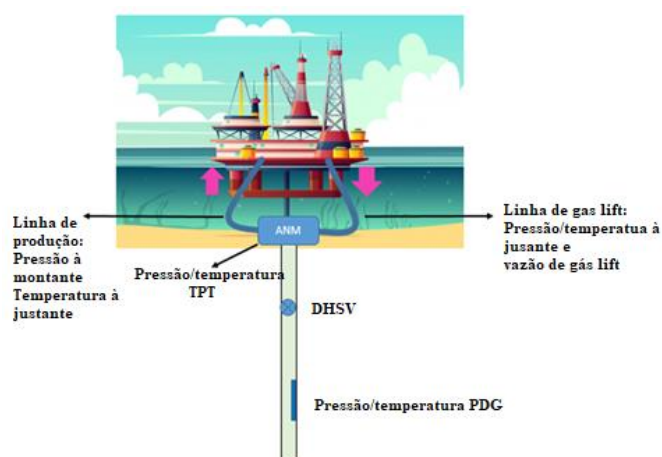


Figura 1- Esquema de uma Unidade Estacionária de Produção

As operações que ocorrem durante a produção de um poço podem ser divididas em dois grupos 1- operações de subsuperfície e 2-operações de superfície. O primeiro grupo engloba operações que ocorrem no solo marinho, onde estão localizados os equipamentos e linhas submersas, como operações de intervenção de poços (*workovers*), operações por cabo (*wireline*), estimulação de poços, elevação artificial por *gas lift* etc. O segundo grupo engloba as operações que ocorrem na UEP (Unidade Estacionária de Produção) como separação, tratamento, armazenamento e transporte de hidrocarbonetos (LAIK, 2018).

A arquitetura submarina de sistemas de produção *offshore*, como mostra a Figura 2, é composta por poços localizados no solo marinho, equipamentos que ligam os poços às linhas de produção, chamados de *manifold* e *risers* que conectam as linhas de produção apoiadas sobre o solo marinho à unidade de produção. Há também linhas destinadas à importação e exportação de hidrocarbonetos (GUO *et al*, 2017).

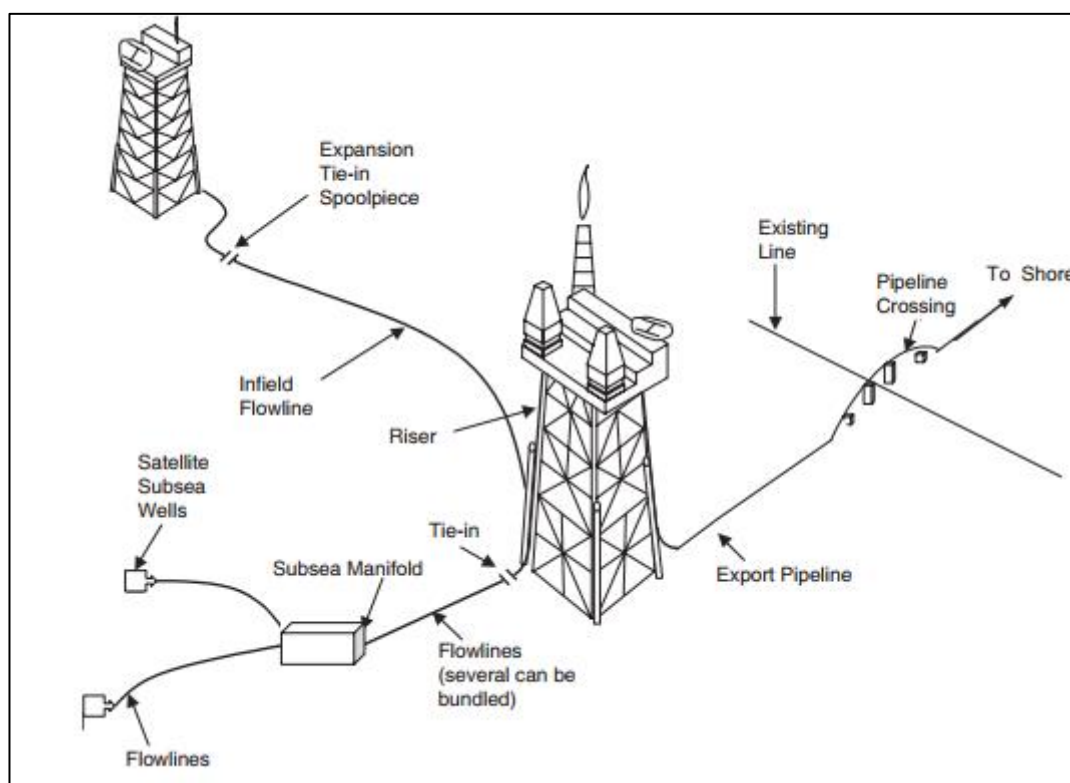


Figura 2- Arquitetura submarina de um sistema de produção *offshore*

Fonte: Guo *et al* (2017).

As unidades de produção podem variar de acordo com a lâmina d'água e o tipo de projeto. No Brasil, há cinco tipos de plataformas em operação: fixa, semissubmersível, FPSO (*Floating Production Storage and Offloading*), FPSO monocoluna e TLWP

(*Tension Leg Wellhead Platform*). As mais comumente encontradas são as fixas, semissubmersíveis e os FPSOs. As unidades fixas, como ilustrado na Figura 2, são fixadas no solo marinho e usadas em lâminas d'água de até 300 metros. O escoamento da produção nesse tipo de plataforma é realizado por meio de oleodutos. As plataformas semissubmersíveis podem ser usadas em grandes lâminas d'água (acima de 2000 metros) possuem um moderno sistema de ancoragem, e podem, também, contar com sistemas de posicionamento dinâmico. Essas plataformas não possuem capacidade de armazenamento da produção, sendo assim, são necessários oleodutos ou navios de armazenamento para posterior descarregamento da produção em terminais. Os FPSOs também podem ser utilizados em lâminas d'água acima de 2000 metros e, diferentemente das semissubmersíveis, possuem alta capacidade de armazenamento. Após essa capacidade ser atingida, é realizada a operação de *offloading*, por meio da qual o óleo que estava armazenado na unidade de produção é transferido para outro navio que será responsável por descarregar a produção (PETROBRAS, 2014).

#### 2.1.1. Sensores

Nesta sessão serão apresentados os principais sensores presentes em sistemas *offshore*, cujos dados serão usados para realização dos experimentos propostos nesta dissertação. Esses são sensores de pressão e temperatura localizados dentro do poço, na árvore de natal, na linha de produção que chega à superfície e na linha de *gas lift* que é conectada na plataforma como pode ser observado na Figura 1.

O PDG (*Permanent Downhole Gauge*) é um sensor de grande importância para a modelagem de poços de petróleo, pois mede a pressão e a temperatura próximas ao fundo do poço (APIO *et al*, 2019). Com a crescente melhoria na confiabilidade desses sensores, eles vêm sendo cada vez mais utilizados em poços *offshore* (GRINGARTEN *et al*, 2003). No entanto, devido às condições mais severas de pressão e temperatura, esses sensores podem ter sua vida útil reduzida. Como o PDG está instalado dentro do poço, em caso de falha não são imediatamente substituídos, pois as operações de intervenção são extremamente custosas. Para continuar tendo acesso aos dados de sistemas produtivos, algumas empresas aumentam o número de sensores para que, caso algum apresente defeito, existam outros funcionando.

Acima da válvula de segurança de subsuperfície (DHSV), está localizada a árvore de natal molhada que acomoda diversas válvulas e sensores de controle e monitoramento do poço. Na árvore de natal estão os sensores T-TPT (*Temperature at the Temperature*

and Pressure Transducer) e P-TPT (*Pressure at the Temperature and Pressure Transducer*). Esses sensores são responsáveis por aferir a pressão e a temperatura do fluido que sai do poço e entra na linha de produção.

Na chegada da linha de produção à UEP são encontradas válvulas *choke* que permitem controlar a vazão do fluido que chega à superfície. A válvula *choke* é usada para controle da produção, ela permite o controle do fluxo de hidrocarbonetos produzidos além de permitir o balanceamento entre a pressão do reservatório e a pressão necessária para escoar a produção até a superfície (TATTERSALL, 2016). No esquema mostrado na Figura 1, há sensor de pressão à montante da válvula *choke* e sensor de temperatura à jusante da válvula.

Ao longo da vida produtiva de um poço de petróleo ocorre queda em sua pressão original. Para manutenção da produção faz-se necessário o uso de ferramentas de elevação artificial. Dentre as técnicas de elevação artificial, destaca-se o *gas lift*, a elevação por injeção de gás no interior da coluna de produção, a partir do espaço anular. A injeção de gás faz com que o fluido escoado se torne menos denso facilitando o escoamento da produção até à superfície. O gás é injetado pela linha de serviço, que é usada também para outras operações. No cenário mostrado na Figura 1, existem sensores de pressão e de temperatura à jusante da válvula *choke* de *gas lift*, que permitem o monitoramento da vazão de injeção.

A arquitetura *subsea*, o sistema de medição e monitoramento, bem como o tipo de unidade de produção podem variar conforme cada projeto. Muitos fatores são importantes para determinar a melhor estrutura para um projeto *offshore*. Além dos desafios técnicos, há também fatores econômicos determinantes para a implementação do projeto. Além disso, como as condições *offshore* são mais severas que em ambientes terrestres, alguns equipamentos têm sua vida útil reduzida e é preciso buscar alternativas para aumentar sua disponibilidade ou implantar mais redundâncias para que o monitoramento ou controle da produção não seja prejudicado.

## 2.2. Garantia de escoamento

Na produção de poços marítimos existem diversos fatores que tornam o escoamento da produção mais complexo. A conjuntura de quilômetros de tubulações dispostas sobre o irregular solo marinho, temperaturas baixas e a presença de diversos equipamentos fazem com que o entendimento do comportamento do fluxo seja primordial

para garantir que a produção de fluidos ocorrerá de forma eficiente (GUDMUNDSSON, 2018).

Nos anos de 1990 a Petrobras usou pela primeira vez o termo “Garantia de Escoamento”, que ficou conhecido internacionalmente como “*Flow Assurance*”, para designar tecnologias e procedimentos que assegurassem a qualidade do escoamento de hidrocarbonetos (LAIK, 2018).

A garantia de escoamento durante a produção em lâminas d’água ultraprofundas foi apontado como o principal desafio técnico que a indústria de petróleo enfrenta. As condições extremas de operação e a formação de sólidos que podem bloquear a linha de produção e gerar grandes impactos financeiros são as principais preocupações.

Portanto, o desafio da produção em campos de petróleo em ambientes marinhos é assegurar o fluxo contínuo de hidrocarbonetos à plataforma. Porém, a baixa temperatura da água do mar em grandes profundidades e as altas pressões hidrostáticas formam um cenário propício para deposição de sólidos como hidratos de gás. O bloqueio da linha de produção é preocupante pois geram altos custos às empresas. Além de interromper a produção é necessário realizar onerosas operações de correção para colocar o poço novamente em produção (LUNA-ORTIZ *et al*, 2008).

Assim, os pontos-chaves da garantia de escoamento são o estudo da deposição de sólidos nas linhas de produção e transporte, tais como asfaltenos, parafinas, hidratos de gás e naftenos, além da formação de depósitos inorgânicos, denominados de incrustação (GUDMUNDSSON, 2018).

Otimizar sistemas de produção de modo a reduzir riscos de problemas de garantia de escoamento pode ser um desafio técnico que aumenta os custos do projeto. Tais riscos podem ser reduzidos através da implementação de sistemas de isolamento térmico, injeção de inibidores químicos e monitoramento extensivo (GUO *et al*, 2013).

Como o objetivo desta pesquisa está relacionado aos problemas de garantia de escoamento como hidratos, incrustação, *slugging* e instabilidade de fluxo esses quatro temas serão abordados com mais detalhes a seguir.

#### 2.2.1. Hidratos de gás natural

Hidratos de gás natural são também chamados de clatratos ou compostos de inclusão. São caracterizados por estruturas que possuem vulgos ou canais formados por moléculas de água que aprisionam pequenas moléculas de hidrocarbonetos como metano, propano e etano (SLOAN; KOH; SUM, 2011). Os hidratos de gás são classificados em

tipos I, II e H. O hidrato tipo H não está relacionado à produção de hidrocarbonetos. Os tipos I e II diferem-se quanto à sua estrutura cristalina. O tipo I ocorre na presença de moléculas menores como metano e etano, ou até mesmo uma mistura de ambos. O tipo II ocorre quando na presença de moléculas maiores como iso-butano ou propano associados a moléculas de metano e etano. Os dois tipos de estruturas podem existir simultaneamente, todavia a estrutura do tipo II é a forma mais estável de hidrato (GLÉNAT, 2015). A Figura 3 mostra a ocorrência de hidrato de gás dentro de uma linha de produção de petróleo.

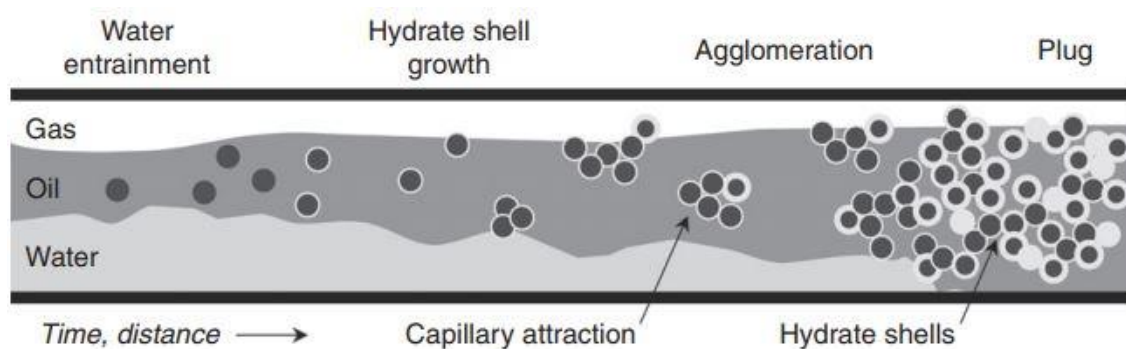


**Figura 3– Plug de hidrato dentro de uma linha de produção**

Fonte: Khana *et al* (2019).

O bloqueio de linhas devido aos hidratos não é um processo imediato, pois a formação de hidrato é caracterizada por fases. A Figura 4 representa o processo de formação de hidratos em linhas onde o fluido dominante é o óleo e há baixa porcentagem de água. No total, quatro etapas são responsáveis pela formação de *plugs* de hidratos: presença de água em emulsão, formação de cristais de hidratos, aglomeração e tamponamento da linha. Primeiro, a água presente na mistura encontra-se dispersa na fase óleo em forma de gotículas. A seguir, conforme as condições e operações adentram a região de formação de hidratos, cristais de hidratos começam a crescer ao redor das

gotículas de água na interface água-óleo. A cristalização das gotículas de água continua da sua superfície externa ao núcleo. Outras gotículas de água livre, que estão próximas às gotículas de água que estão passando pelo processo de hidratação, ficam à mercê da ação de forças capilares de atração que desencadeiam o processo de aglomeração. Com o passar do tempo, a aglomeração dos cristais de hidratos continua e, caso nenhuma ação corretiva seja tomada, a linha será completamente obstruída, o que pode ocasionar a parada completa da produção, fator que exige intervenções onerosas para retomada do processo produtivo (SLOAN; KOH; SUM, 2011).



**Figura 4- Processo de formação de hidratos**

Fonte: Sloan, Koh e Sum (2011).

Os hidratos de gás são formados em condições de baixas temperaturas (menor que 20°C) e pressões moderadas a altas (acima de 10bar). Tais condições são frequentemente encontradas em sistemas de produção de petróleo *offshore* (LACHANCE; KEINATH, 2015). Paradas de produção que acarretam a diminuição da temperatura na linha de produção e injeção inadequada de inibidores em sistemas que possuem coexistência de gás e água livre também são situações que podem promover a formação de hidratos. O processo de formação também é favorecido em regiões onde a velocidade do fluxo é alta, por esta razão válvulas *choke* são suscetíveis à formação de hidratos (OSOKOGWU; AJIENKA, 2010).

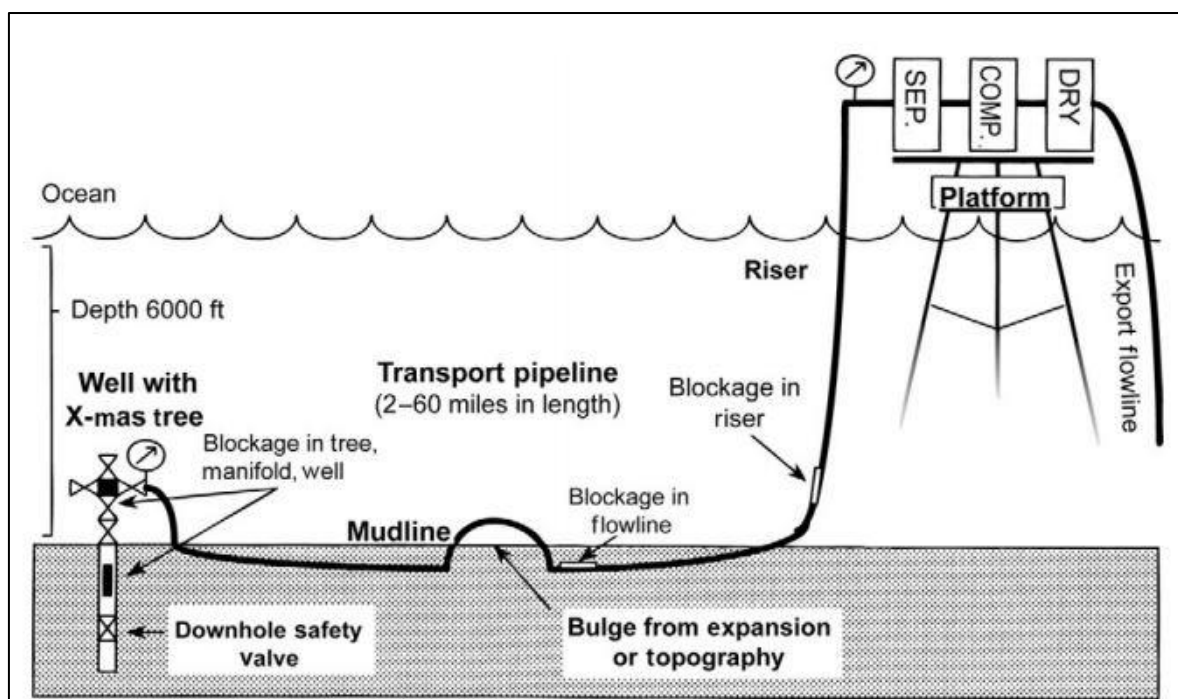
É comum que linhas de produção possuam isolamento térmico para manter a temperatura do escoamento a mais alta possível, até sua chegada à plataforma. Em sistemas de produção onde o óleo é a fase predominante a perda de calor ao longo das linhas de produção é muito menor, quando comparado a sistemas onde o gás é o fluido dominante. Por esse motivo, em sistemas onde a produção é predominantemente de gás, são necessárias estratégias adicionais como injeção de inibidores para diminuir o risco de formação de hidratos (SLOAN; KOH; SUM, 2011).

Considerando uma arquitetura submarina simplificada, a Figura 5 mostra locais com maior risco de formação de hidratos de gás. A figura apresenta um poço produtor onde o fluido é produzido no reservatório, passa pela árvore de natal (*X-mas tree*) e/ou manifold e percorre a linha de produção sinalizada pela linha preta. Como o solo marinho é irregular, a figura também representa variações de topografia que podem existir no fundo do mar. Em seguida, é ilustrada a linha que fica apoiada no solo marinho passando para o trecho vertical (*riser*) que se conecta à plataforma, onde ocorre o processo de separação de fluidos, compressão de gás e retirada da água presente no gás para posterior exportação. Os pontos de risco de formação de hidratos estão na saída do poço, árvore de natal e *manifold*, onde ocorre queda da temperatura; no entanto, caso haja injeção de inibidores pela árvore de natal, é baixa a probabilidade de formação de hidratos nesses pontos.

A formação de hidratos também pode ocorrer em depressões originadas por irregularidades do solo marinho e na base dos (*risers*), lugares preferenciais de acúmulo de água (fração mais pesada do fluido). Nesses lugares são criados ambientes propícios para que moléculas de gás entrem em contato direto com a água em condições que favorecem a hidratação da linha.

Uma outra situação que favorece a formação de hidratos é a parada de produção (*shutdown*). Nestas, o fluido contido nas linhas entra em equilíbrio térmico com o ambiente marinho e se separa em fases por efeitos gravitacionais, permitindo o surgimento de pontos de acumulação de água. Portanto, em casos de parada de produção nos quais condições favoráveis à formação de hidratos são induzidas, é necessário implementar medidas de prevenção para retomar o processo de produção. Nesses casos, muitas empresas optam por aumentar a injeção de inibidores e circular fluidos nas linhas para aumentar a temperatura das mesmas (SLOAN; KOH; SUM, 2011).





**Figura 5- Pontos de maior risco de formação de hidratos em sistemas offshore**

Fonte: Sloan, Koh e Sum (2011).

Sendo um problema comum na indústria de petróleo, muitas empresas já incluem em seus projetos medidas para combater a formação de hidratos. O procedimento mais comum adotado pela indústria é manter a produção fora das condições propícias à formação de hidratos. Produzir fora da zona de formação de hidratos requer que as linhas de produção não percam calor a ponto de atingirem a temperatura de formação de hidrato. Para tanto são implementados sistemas de isolamento térmico, ou até mesmo, aquecimento das linhas. Outra forma de impedir a formação de hidratos é alterando a região de formação de hidratos para temperaturas menores do que a temperatura de operação por meio da injeção de inibidores termodinâmicos como metanol e MEG (Mono Etileno Glicol). No entanto, essa solução apresenta grande impacto financeiro ao projeto, além de limitações técnicas e ambientais (GLÉNAT, 2015).

Durante a produção de petróleo, caso as medidas de injeção de inibidores e de isolamento térmico não sejam eficientes, ainda há a possibilidade de formação de hidratos. A formação de hidratos é detectada por meio de indícios encontrados durante operações de limpeza da linha (*pigging*), mudança na vazão, deposição de sólidos no separador, aumento da queda de pressão no sistema e mediante métodos acústicos (OSOKOGWU; AJIENKA, 2010).

A formação de hidratos nas linhas de produção pode levar à obstrução parcial ou completa da linha. A maneira mais comum de remover os hidratos é por meio de sua dissociação, que pode ser obtida como resultado de métodos hidráulicos, químicos e térmicos. O método hidráulico ocasiona a despressurização da linha onde se formou o hidrato e é o mais aplicado dentre os três métodos, sendo comumente usado em caso de obstrução total da linha. A despressurização, redução da pressão da linha, provoca a dissociação do hidrato por mudar as condições de pressão e temperatura para um estado em que não ocorre a formação de hidratos. O método químico é mais usado para obstruções parciais e consiste na injeção de inibidores termodinâmicos que diminuem a temperatura de formação, fazendo com que as condições de operação não sejam propícias para formação de hidratos. Por último, o método térmico envolve o aquecimento direto da região onde o hidrato está formado, promovendo assim sua dissociação (OSOKOGWU; AJIENKA, 2010).

#### 2.2.2. Incrustação

Dentre os problemas que podem ocorrer durante a produção de hidrocarbonetos está a incrustação resultante da deposição na parede interna das tubulações de determinados compostos químicos e material inorgânico, presentes na água de produção e em outros fluidos produzidos.

A formação de incrustação está diretamente ligada à presença da fase aquosa na mistura de hidrocarbonetos. Existem quatro tipos de água que são coproduzidas com o petróleo: água de formação, água condensada, água de injeção e água proveniente de aquíferos e, em conjunto, são denominadas água de produção. A água de formação é a água presente no espaço poroso do reservatório de petróleo. A água condensada é oriunda do vapor de água presente na fase gasosa. A água de injeção é a água que foi injetada no reservatório, por meio de um poço injetor, para suporte de pressão e que, eventualmente, chega ao poço produtor após percorrer o reservatório. Por último, a água de aquífero é proveniente de aquíferos localizados abaixo da camada de óleo e gás que promovem suporte de pressão ao reservatório (GUDMUNDSSON, 2018).

O fluido de produção apresenta uma variedade de substâncias dissolvidas que são afetadas pelas quedas de pressão e mudanças de temperatura que ocorrem ao longo do escoamento. Com essas mudanças de pressão e temperatura, alguns compostos químicos podem se tornar mais estáveis no estado sólido e começar um processo de precipitação. Métodos de recuperação que visam aumentar o volume de petróleo

recuperado através da injeção de água no reservatório também podem favorecer o processo de incrustação e possível dano à formação rochosa (OLAJIRE, 2015).

Dentre as substâncias dissolvidas na água de produção, destacam-se diversos minerais que provocam a presença de íons. Normalmente, esses íons incluem cálcio ( $\text{Ca}^{2+}$ ), bário ( $\text{Ba}^{2+}$ ), estrôncio ( $\text{Sr}^{2+}$ ), carbonatos ( $\text{CO}_3^{2-}$ ) e sulfatos ( $\text{SO}_4^{2-}$ ). Nos campos de petróleo, a incrustação costuma consistir em carbonato de cálcio, sulfato de cálcio, sulfato de bário, sulfato de estrôncio e carbonato de estrôncio (OLIJERA, 2015). A Figura 6 mostra como a precipitação de sólidos pode acarretar a diminuição significativa do diâmetro disponível para o escoamento da produção.



**Figura 6- Exemplo de incrustação em uma linha de produção em terra**

Fonte: Olajire (2015).

A deposição de sólidos que origina o processo de incrustação envolve diversos fatores químicos e termodinâmicos, dos quais os principais são a concentração de certas substâncias na água de produção, a velocidade das reações químicas envolvidas, as condições de pressão e temperatura, a incompatibilidade entre fluidos na mistura e a alcalinidade do meio. Quando as substâncias dissolvidas na água de produção ultrapassam o limite de solubilidade, dá-se início ao processo de precipitação. A depender da cinética

da reação de precipitação que ocorre, é possível determinar se haverá deposição local (cinética rápida) ou se a deposição ocorrerá de forma gradual, a partir da região onde o limite de solubilidade foi alcançado (cinética lenta). As condições de pressão e temperatura afetam a solubilidade e o produto de solubilidade da mistura. A presença de substâncias incompatíveis na mistura pode favorecer a precipitação de sólidos mais estáveis. A alcalinidade da mistura representa a sua capacidade de neutralizar ácidos, tendo como subproduto a formação de sais (GUDMUNDSSON, 2018).

Durante a produção de hidrocarbonetos, incrustações podem ser formadas também no reservatório, próximo ao fundo do poço, em bombas submersas, tubos de produção e válvulas. Na etapa de perfuração e completação também é possível que haja precipitação de sólidos, principalmente quando o fluido de perfuração utilizado é à base de água. Por questões de segurança, a maior preocupação dos operadores é a incrustação em válvulas, principalmente válvulas de segurança de subsuperfície, que agem como barreira de segurança, caso algum problema ocorra na plataforma. Outro ponto importante é a incrustação de substâncias radioativas como sulfato de rádio ( $\text{RaSO}_4$ ) que forma depósitos junto com outras substâncias maiores, como sulfatos de bário e cálcio. A presença de substâncias radiotivas apresenta riscos para as pessoas que manuseiam válvulas, além de necessitar de tratamento especial para serem descartadas (GUDMUNDSSON, 2018).

Gerenciar a formação de incrustações em ambientes *offshore* não é trivial. Devido às condições extremas e aos altos custos envolvidos, estabelecer um plano de prevenção contra as incrustações se faz essencial. Para fins de injeção de água no reservatório, a água do mar, presente em abundância, se faz muito útil. No entanto, a água do mar apresenta elevada concentração de sulfatos dissolvidos, tais sulfatos potencializam a formação de incrustação dentro do reservatório, próximo ao fundo do poço. De modo a diminuir a concentração de sulfatos, muitas plataformas possuem uma unidade para remoção de sulfatos. Após a remoção dos sulfatos, é menos provável que a água do mar injetada no reservatório dê origem às incrustações por sulfatos. Outra medida que reduz o potencial de formação de incrustações no sistema é a injeção de inibidores químicos, num procedimento chamado *squeeze*, no qual os inibidores são injetados dentro da coluna de produção, próximos ao fundo do poço. O inibidor tem como função controlar, retardar e diminuir a formação de incrustações, além disso é mais econômico e efetivo na prevenção de incrustação (OLIJERA, 2015).

Nos casos para os quais a linha ou válvulas já se encontram incrustadas, há dois métodos de remoção desses depósitos: mecânico e químico. O método mecânico consiste em remover a incrustação através do uso de brocas. Para tanto, é necessário alugar sondas o que torna o método muito caro. O método mecânico não é eficaz em caso de incrustações localizadas próximas ao fundo do poço ou na formação rochosa. O método químico costuma ser mais barato que o método mecânico, e consiste na aplicação de ácidos e outras substâncias capazes de dissolver os minerais presentes na incrustação. A acidificação é largamente utilizada e também apresenta a vantagem de promover a estimulação do poço ao remover obstáculos ao fluxo de fluido, diminuindo o dano à formação e, assim, aumentando a produtividade do poço (OLIJERA, 2015).

### 2.2.3. Instabilidade de fluxo

O escoamento de dois ou mais fluidos ao longo de uma tubulação é muito mais complexo do que o escoamento monofásico. As fases presentes tendem a se separar devido às suas diferenças em densidade. Também devido à diferença de densidade e de viscosidade entre as fases, a tensão de cisalhamento imposta pela superfície interna da tubulação é distinta para cada fluido escoado. A fase gasosa, por sua alta compressibilidade, em trechos onde há redução da pressão e consequente expansão volumétrica, apresenta maior velocidade que a fase líquida. Portanto, é comum existir diferença entre a velocidade da fase líquida e da fase gasosa. A diferença entre a velocidade das fases depende da configuração espacial da tubulação e do sentido do fluxo. Em escoamentos verticais cujo sentido do fluxo é ascendente (*upstream*), a velocidade do gás, pela redução da pressão e pelo processo de expansão do mesmo, é maior que a velocidade do líquido. Por outro lado, em escoamentos de cima para baixo (*downtream*), o efeito gravitacional favorece a aceleração da fase líquida, sendo assim, nesse tipo de escoamento, a fase líquida costuma apresentar maior velocidade (BRILL; MUKHERJEE, 1999).

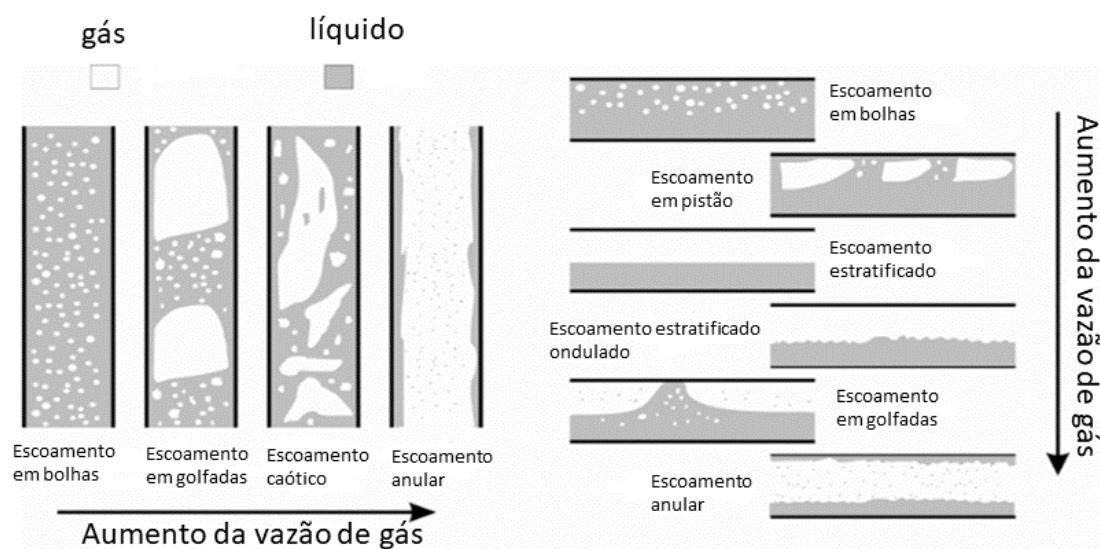
A produção de petróleo engloba o escoamento multifásico, onde óleo, água, gás e possivelmente algumas partículas sólidas são produzidas. O conhecimento acerca de padrões de escoamento e mecânica dos fluidos é importante para modelar, controlar e mensurar as variáveis envolvidas no processo de escoamento. Além disso, é importante para descrever mudanças que ocorrem no escoamento em estruturas verticais, horizontais e inclinadas (TAN *et al*, 2015).

O escoamento multifásico dentro de uma tubulação, linha de produção ou *riser* pode assumir diferentes distribuições físicas, que são conhecidas como padrões de escoamento. O padrão de escoamento vai depender da intensidade das forças que atuam sobre o escoamento, como empuxo, turbulência, inércia e forças relacionadas à tensão superficial. Essas forças variam de acordo a vazão de escoamento, diâmetro do tubo, ângulo de inclinação do tubo e propriedade dos fluidos envolvidos. Devido à variação de temperatura e, principalmente, de pressão, ao longo do escoamento, podem existir diversos padrões de escoamento num poço (BRILL; MUKHERJEE, 1999).

Em escoamentos verticais de fluxo ascendente, como os que ocorrem dentro de *risers* e tubos verticais, há quatro padrões: escoamento em bolhas (*bubbly flow*), escoamento em golfadas (*slug flow*), escoamento intermitente (*churn flow*) e anular (*annular flow*), como mostra a primeira parte da Figura 7. O padrão de escoamento em bolhas é caracterizado por bolhas de gás dispersas uniformemente na fase líquida (fase contínua). O escoamento em golfadas é representado por uma série de bolhas grandes, chamadas de bolhas de Taylor, que possuem entre si uma secção de líquido contendo bolhas menores dispersas. A seguir, o escoamento intermitente, apresenta um fluxo caótico onde as bolhas de Taylor e as secções de líquido encontram-se distorcidas e nenhuma das fases aparenta ser contínua. Por último, o escoamento anular apresenta a fase gasosa como contínua, enquanto a fase líquida está ou contida em gotículas dispersas no interior da fase contínua ou, na parede da tubulação, na forma de uma película de líquido (BRILL; MUKHERJEE, 1999).

Em sistemas de produção de petróleo também é comum a presença de trechos de escoamento horizontal como linhas de produção, importação e exportação. Na segunda parte da Figura 7 é possível observar os regimes de escoamento que ocorrem em trechos horizontais. Devido à influência de forças gravitacionais, o escoamento horizontal apresenta algumas distinções do escoamento vertical ascendente. O escoamento em bolhas é caracterizado por bolhas dispersas na parte superior do fluido, devido à diferença de densidade entre as fases. Com a coalescência das bolhas que antes encontravam-se dispersas na parte superior, dá-se início ao regime de bolhas alongadas (*plug flow*). Com o processo de coalescência ainda em ação, a fase gasosa passa apresentar uma camada contínua, caracterizando o escoamento estratificado (*stratified flow*). Como o gás apresenta menor densidade e viscosidade que a fase líquida sua velocidade é maior. Sendo assim, a diferença entre as velocidades das fases é suficiente para gerar ondas na interface gás-líquido dando origem ao escoamento ondulado (*wavy flow*). O escoamento por

golfadas (*slug flow*) diferentemente do que ocorre em tubos verticais, ocorre quando as ondas geradas na interface gás-líquido crescem a ponto de quebrar a continuidade da fase gasosa e alcançar a parede do tubo. Por último, o escoamento anular, assim como em tubos verticais, apresenta a fase gasosa como fase contínua com gotas de água dispersas em seu interior e uma película de líquido ao longo da parede do tubo. No entanto, devido a forças gravitacionais, a película de líquido concentra maior espessura na parte inferior do tubo (HOLLAND; BRAGG, 1995).



**Figura 7 -Padrões de escoamento em tubulações verticais, de fluxo ascendente, e horizontais**

Fonte: Da Silva (2014).

Problemas de instabilidade no fluxo do escoamento da produção podem acarretar perdas significativas de produtividade. Em caso de instabilidade de fluxo de alta amplitude, como é o caso do *slug*, a perda de produção pode passar de 20%, além de aumentar os custos operacionais. Alguns fatores que induzem instabilidade de fluxo são: vazão de óleo, vazão de *gas lift* injetado, oscilações na pressão, redução no diâmetro disponível das válvulas de produção e na cabeça do poço (*wellhead*), além da mudança na profundidade de injeção do *gas lift* dentro do poço (GUERRERO-SARABIA; FAIRUZOV, 2013).

#### 2.2.4. *Slugging*

A eficiência do sistema de processamento de uma unidade de produção de petróleo é influenciada pela variação da composição e pelas condições de estabilidade do fluxo de fluido que chega até à unidade. Caso o fluxo esteja instável em sua chegada à plataforma,

o sistema de processamento pode sofrer impactos significantes que resultem na parada de todo o sistema de produção. Na produção *offshore*, quanto maior a lâmina d'água alcançada, mais energia é necessária para transportar o fluido produzido, do leito marinho até à plataforma.

Quando a velocidade do gás e do líquido não são suficientemente altas para manter de forma contínua o transporte do fluido, dá-se origem a um regime instável de intermitência na chegada dos fluidos à planta de processamento. Como o escoamento do fluido não se dá numa fase contínua, o líquido passa a se acumular na base do *riser*, formando uma coluna de líquido chamada de *slug* líquido (Fig. 8-a). Essa crescente coluna de líquido passa, então, a bloquear a passagem de gás, que se acumula em um bolsão, abaixo da coluna de líquido. A coluna de líquido pode aumentar até atingir a altura total do *riser* e entrar na planta de processamento ou, então, até que o bolsão de gás à montante tenha pressão suficiente para deslocá-la até a superfície (Fig. 8-b). Ao chegar nos equipamentos de superfície, a coluna de líquido apresenta alta velocidade e nenhum ou pouco gás. O separador de produção destinado a receber a mistura de água, óleo e gás recebe, assim, um grande volume de líquido a alta velocidade, fato que ocasiona rápido alcance da capacidade máxima de processamento; ademais a ausência de gás no fluido pode ocasionar problema nos compressores da instalação. Na sequência, o bolsão de gás que se acumulava abaixo da coluna de líquido chega à superfície (Fig. 8-c). Devido à ausência de líquido, o bolsão de gás pode gerar o desligamento do separador pois este possui restrições quanto à quantidade mínima de líquido necessária para seu funcionamento. Com o desligamento do separador, toda a produção é afetada e até os equipamentos submarinos precisam ser desligados (Fig. 8-d). Após a chegada da coluna de líquido e do bolsão de gás à plataforma, um novo ciclo de formação de *slug* começa. A Figura 8 apresenta em etapas a formação de *slug* como descrita anteriormente (GUO *et al*, 2013).

Resumidamente, o *slug* severo é indesejado pois gera: grandes flutuações no escoamento de líquido e gás, transbordamento no separador, impacta de forma considerável a pressão dos fluidos que chegam à superfície, vibração das tubulações e intensifica o processo de corrosão. Sendo assim, devem ser monitorados os parâmetros de escoamento no trecho onde há maior risco de ocorrer *slugging* para que o escoamento da produção não seja prejudicado (XU *et al*, 2019).



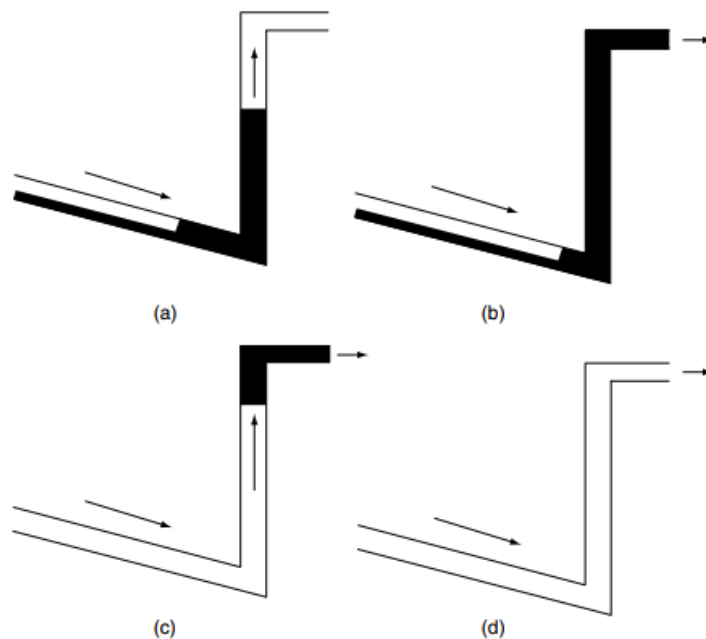


Figura 8- Esquema de formação de slug.

Fonte: Guo *et al* (2013).

Os regimes de escoamento costumam ser determinados a partir de experimentos combinando observação visual e sinais de pressão do sistema. Na produção *offshore*, a lâmina d'água difere para cada área explorada. Além disso, é possível que a estrutura do *riser* também seja distinta. Sendo assim, é difícil estabelecer padrões de escoamento que se adequem para cada configuração de sistema *offshore*. Na literatura, há autores que estabelecem três ou mais tipos de *slug* severo (TIN, 1991 ; ZOU; GUO; XIE, 2017) outros estabelecem apenas dois tipos (SCHMIDT; BRILL; BEGGS, 1980; YE; GUO, 2013) a depender dos tipos de fluidos envolvidos, da configuração e do comprimento do *riser*. Nesse sentido, avanços no desenvolvimento de modelos capazes de generalizar os tipos de regimes de escoamento para comprimentos de *risers* cada vez maiores vêm sendo implementados, alinhados a técnicas de aprendizado de máquina (XU *et al*, 2019).

O *slug* severo diferencia-se do *slug* convencional pela amplitude de flutuação da pressão. Normalmente, o *slug* severo pode ter comprimento equivalente ou superior à altura do *riser* (JANSEN; SHOHAM; TAITEL, 1996). A ocorrência de *slugging* durante a produção de petróleo já é levada em conta durante a modelagem do sistema. Parâmetros que estabelecem critérios de formação de *slug* como velocidade superficial do gás, taxa de acúmulo de gás na base do *riser*, taxa de aumento da pressão do gás e pressão

hidrostática, são controlados de modo a diminuir o risco de *slugging* nos equipamentos e tubulações (PEDERSEN; DURDEVIC; YANG, 2017).

Os métodos de eliminação de *slug* consistem em: mudança no formato do *riser*, imposição de restrições ao longo do escoamento, condicionador de fluxo, *slug catcher*, homogeneizador e *self-gas lifting*. A mudança do formato do *riser* para estruturas não retas tem como objetivo reduzir a região de formação de *slug*, impedindo a acumulação de líquido na base do *riser*. A imposição de elementos que reduzem o diâmetro local da tubulação e restringem o escoamento próximo à base do *riser* faz com os fluidos sejam misturados novamente, criando um regime turbulento e proporcionando uma perda de carga local. Os condicionadores de fluxo também são elementos que, ao serem inseridos num trecho de tubulação, retificam o fluxo e neutralizam irregularidades no perfil do escoamento. O *slug catcher* é um dos métodos mais aplicados, este equipamento consiste em filtrar e receber o grande volume de líquido e gás antes que esses cheguem ao separador de produção. O *slug catcher* funciona como uma etapa de pré-separação na qual a variação do volume de líquido e gás, bem como a flutuação na pressão, não causam danos ao equipamento. O homogeneizador atua transformando a mistura líquido-gás em uma mistura homogênea, eliminando o fluxo multifásico que ocasiona a formação de *slug*. Por último, o método de *self-gas lifting* promove a redução do ciclo de formação do *slug*, reduzindo assim a flutuação da pressão. A injeção de gás no trecho próximo à base do *riser* torna a coluna de líquido gaseificada e mais leve, portanto, a pressão necessária para empurrar essa coluna até a superfície é menor (PEDERSEN; DURDEVIC; YANG, 2017).

# Capítulo 3

## Revisão teórica – Aprendizado de máquina

Neste capítulo serão apresentados conceitos relacionados à área de aprendizado de máquina e inteligência artificial. O conteúdo apresentado a seguir é fundamental para a compreensão dos algoritmos aplicados e dos métodos de avaliação dos modelos gerados neste trabalho.

### 3.1. Mineração de dados e aprendizado de máquina

Especialistas acreditam que padrões incidentes em dados podem ser identificados, validados e usados na predição de eventos futuros de forma automatizada. Com o crescente número de banco de dados disponíveis e também a crescente capacidade de armazenamento de dados, a mineração de dados se torna uma potencial técnica para elucidar e identificar padrões contidos nos dados. Essa técnica é definida como um processo de identificação de padrões em dados de forma automática ou semiautomática. Sendo assim, analisar de maneira inteligente os dados é muito importante para empresas e indústrias, uma vez que pode gerar vantagens competitivas.

Nesse sentido, são de interesse as técnicas capazes de extrair padrões estruturais nos dados disponíveis e utilizá-los como ferramentas para entender os dados e fazer predições a partir deles. O aprendizado de máquina consiste na aquisição de conhecimento a partir da mineração dos dados e em utilizar desse conhecimento em conjuntos de novos dados (WITTEN; FRANK, 2005).

A formação de conhecimento a partir da mineração de dados pode ajudar solucionar problemas de organizações mundias, indústrias e indivíduos. A Figura 9 apresenta o processo de descobrimento de conhecimento (*knowledge discovery*). Esse processo é definido como a extração de informação implícita e útil, antes desconhecida, a partir de dados (BRAMER, 2016). Como observado na Figura 9, a mineração de dados consiste em apenas uma etapa desse processo.

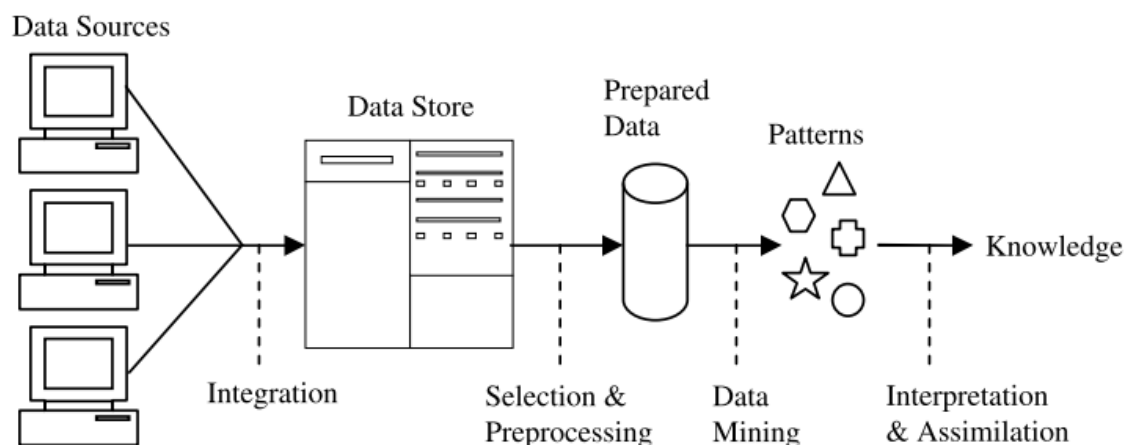


Figura 9-Processo de descobrimento de conhecimento (Knowledge discovery)

Fonte: Bramer (2016).

Os dados usados no processo podem ser oriundos de fontes distintas. No caso de sistemas de produção de petróleo, podem ser oriundos de diversos sensores, de sistemas de monitoramento ou de dados provenientes de análises humanas como relatórios preenchidos por operadores. Então, esses dados são integrados e permanecem armazenados em um diretório. A seguir, parte dos dados são selecionados e preparados para servir como *input* para algoritmos de mineração de dados. A partir da mineração, são identificados padrões e regras, que ao serem interpretados ou assimilados, podem se tornar uma fonte nova e útil de conhecimento (BRAMER, 2016).

### 3.2. Aprendizado supervisionado e não supervisionado

A aprendizagem no contexto de inteligência artificial se dá quando é apresentado ao algoritmo um conjunto de dados limitados e, a partir disso, o modelo construído pelo algoritmo fornece uma descrição concisa dos dados. Se os dados fornecidos representam um conjunto de dados de entrada e saída, com atributos chaves ou alvos, uma boa descrição dos dados é uma função que produz uma saída condizente com os padrões identificados, quando, para ela, for apresentado um dado de entrada. Por outro lado, há casos em que os dados apresentados ao algoritmo não possuem um atributo alvo. Neste caso, a descrição dos dados pode ser um conjunto de *clusters* ou distribuições de densidade mostrando como um dado de entrada pode vir a se comportar no futuro (HERBRICH, 2002).

Um conjunto de dados pode ser considerado classificado ou não classificado, a depender dessa informação o tratamento dos dados é completamente diferente. Dados

classificados ou marcados possuem um atributo chave, onde o objetivo é utilizar os dados obtidos à priori para prever o valor do atributo chave de dados novos, ainda não utilizados no modelo. Se o atributo chave é categórico, ou seja, está dividido em categorias, a tarefa realizada pelo modelo é de classificação. Caso o atributo chave seja numérico, a tarefa realizada é uma regressão. Dados não classificados não possuem atributo chave e o objetivo principal ao se trabalhar com este tipo de dado é extrair o máximo de informação possível. Em relação aos tipos de aprendizado, dados classificados são utilizados em técnicas de aprendizado supervisionado, enquanto dados não classificados são a base do aprendizado não supervisionado (BRAMER, 2016).

Classificação é a aplicação mais comum dentro da mineração de dados. Regressão também é uma forma de classificação, na qual o atributo alvo é um valor numérico e não uma classe (BRAMER, 2016). Em um problema de aprendizado supervisionado é dada uma amostra de treino contendo dados de entrada e saída. A tarefa do algoritmo é estabelecer uma função que estabeleça o resultado de saída para um dado de entrada novo, tendo como objetivo o valor ou classe do atributo chave desse novo dado (HERBRICH, 2002).

Quando o conjunto de dados apresentado a um algoritmo não possui uma marcação que possa ser usada para representar concisamente os dados é necessária uma abordagem diferente. Em objetos não marcados, faz sentido identificar similaridades entre os dados, de modo que eles possam ser dispostos de acordo com suas semelhanças. Esse é o conceito em que se baseiam os algoritmos de agrupamento (*clustering*), os quais basicamente organizam os dados de acordo com um número fixo de grupos (*clusters*), dessa forma é possível visualizar melhor o conjunto de dados e extrair informações úteis (HERBRICH, 2002).

### 3.3. Métodos de partição do conjunto de dados

O objetivo de um algoritmo de classificação é possibilitar a classificação de elementos novos, não processados previamente pelo algoritmo. Para mensurar a habilidade de um modelo em classificar novas instâncias é preciso estabelecer métricas e parâmetros de comparação. Há três métodos popularmente usados para realizar essa tarefa; a divisão em conjunto de treino e teste, a validação cruzada usando *k-folds* e a validação cruzada *N-folds*.

A validação cruzada usando  $k$ -folds é explicada na figura 10. Esse método é utilizado em conjuntos de dados considerados pequenos. Ele consiste em dividir o conjunto de dados em  $k$  partes e, a partir dessa divisão, o modelo é construído usando  $k-1$  partes como conjunto de treino e a última  $k$  parte é usada para testar o modelo (BRAMER, 2016).

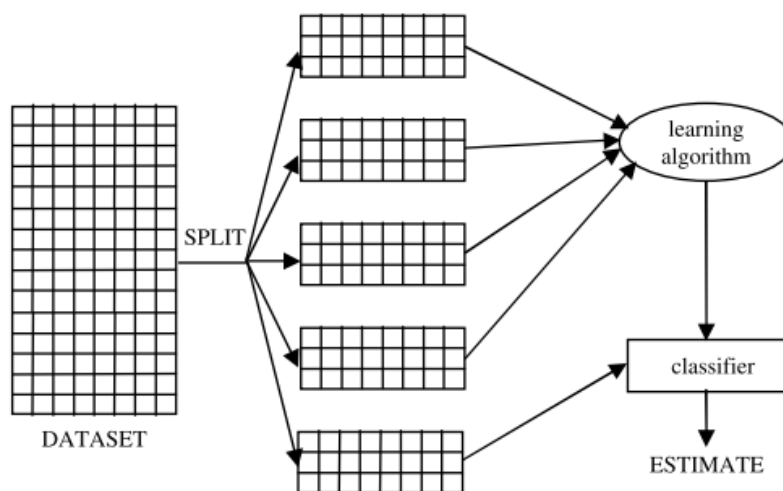


Figura 10- Esquema de validação cruzada usando  $k$ - folds

Fonte: Bramer (2016).

Por outro lado, a validação cruzada com  $N$ -folds é um caso extremo da validação cruzada exposta na figura 10. Neste caso, dado um conjunto contendo  $N$  instâncias, serão gerados  $N-1$  modelos de classificação. Cada um dos  $N-1$  classificadores serão testados com a única instância que não for utilizada previamente. A acurácia do algoritmo de classificação, utilizando esse método, considera o número total de instâncias classificadas corretamente e o número total de instâncias  $N$ . Como esse método exige maior esforço computacional, ele não é recomendado para grandes volumes de dados (BRAMER, 2016).

O método de partição dos dados de treino e teste, apresentado na Figura 11, consiste em dividir o conjunto original em duas partes; uma usada para treinar e construir o modelo de classificação e outro para testar o modelo. Com isso, o conjunto de teste fornece novas instâncias que não foram usadas durante o treinamento. A acurácia do classificador é estimada através da quantidade de instâncias classificadas corretamente e o número de instâncias total no conjunto de teste.

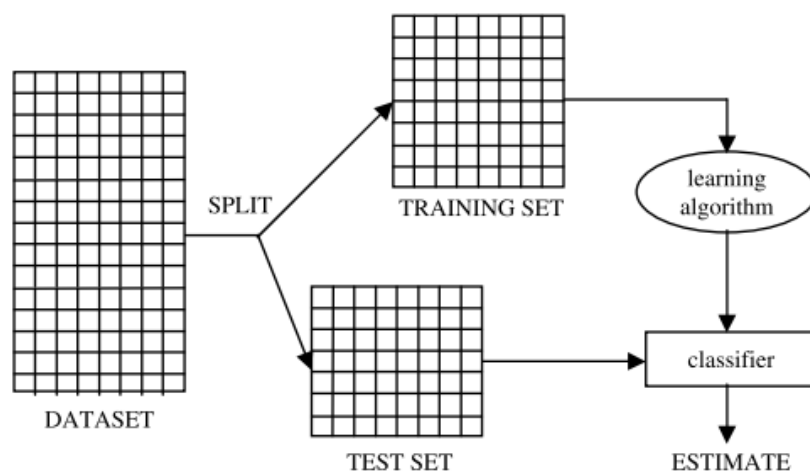


Figura 11- Esquema de partição de conjunto de treino e teste.

Fonte: Bramer (2016).

O método de separação em conjunto de treino e teste foi escolhido dentre os anteriores pela sua facilidade de aplicação e pelas características dos dados usados nesta pesquisa. O banco de dados usado contempla grandes volumes de dados e as classes presentes nele não estão balanceadas. O desbalanceamento entre as classes significa que apenas uma pequena parte dos dados apresenta as classes “normal” e “anormal”, a grande maioria das instâncias são referentes à classe de transição. Sendo assim, segundo Martloff (2017), ao separar aleatoriamente as amostras de menor tamanho, como nos métodos de validação cruzada, é possível que alguma classe deixe de ser selecionada.

### 3.4. Algoritmos de classificação

Problemas de classificação referem-se a tomar um dado de entrada e tentar prever a qual classe ou atributo chave este dado pertence. Em alguns casos, também é de interesse entender melhor as variáveis que descrevem um problema. Por exemplo, se é de interesse de um hospital identificar quais fatores influenciam na sobrevivência de pacientes internados, seria importante analisar quais fatores estão relacionados à classe de pessoas sobreviventes. Nesse caso, mais importante do que classificar se um paciente sobreviveu ou não, é identificar ou descrever que fatores foram relevantes para esse resultado. Sendo assim, para problemas de classificação é possível haver dois objetivos a serem alcançados: predição e descrição (MATLOFF, 2017).

Ferramenta de aprendizados de máquinas podem ser úteis para a visualização de informações a cerca de um conjunto de dados. Durante a resolução de problemas de

classificação é importante definir parâmetros e variáveis que influenciam de maneira significativa o comportamento dos dados. Assim, é possível melhorar, descritivamente, a compreensão do problema a ser abordado.

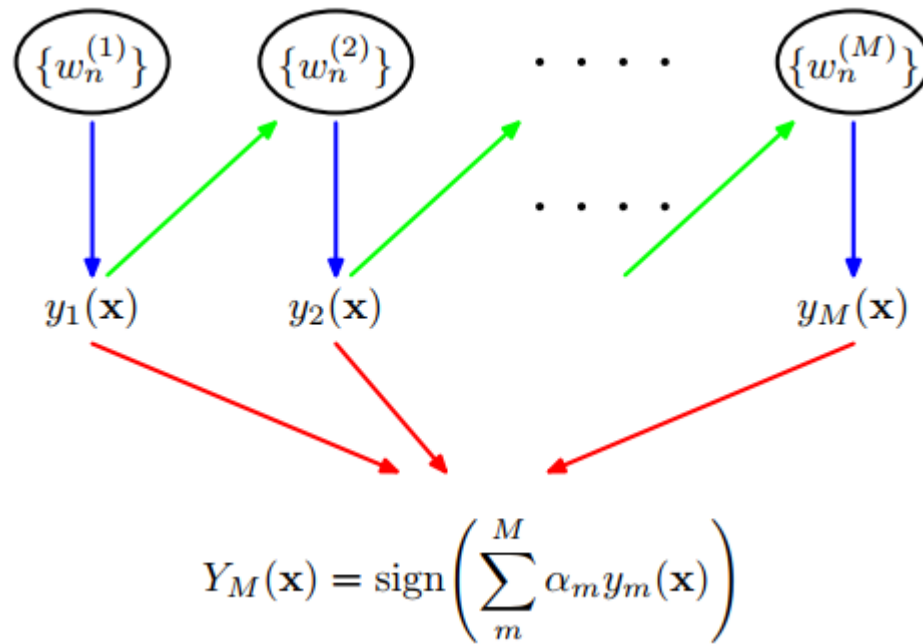
A seguir, serão fornecidos mais detalhes sobre os algoritmos utilizados no conjunto de dados para classificar dados referentes à formação de hidrato, incrustação, instabilidade de fluxo e *slugging*.

#### 3.4.1. Adaboost

*Boosting* é uma técnica de classificação que vem sendo muito aplicada na área de aprendizado de máquina. Sua essência é encontrar funções de regressão e classificação através de um processo iterativo de refino e ponderamento de parâmetros. Sumariamente, os algoritmos de *boosting* ajustam continuamente o modelo criado, reavaliando o ponderamento dos parâmetros a cada iteração (MARTLOFF, 2017). Dentre os algoritmos de *boosting*, o mais amplamente utilizado é o Adaboost, cujo nome representa a abreviação para *adaptive boosting*, criado em 1996 por Freund e Schapire (NIELSEN, 2016).

O *adaboost* realiza um processo de aperfeiçoamento nos classificadores. A Figura 12 representa o esquema de processamento do algoritmo. Primeiro, começando pela esquerda, as instâncias do conjunto de treino são ponderadas com pesos denominados  $w_n$  (balões em negrito). Em seguida, as instâncias são usadas para treinar o classificador denominado pela letra  $y$  (seta azul). As instâncias que forem classificadas erroneamente recebem então maior peso na segunda iteração (seta verde). Em suma, a cada iteração as instâncias com classificação errada recebem maior peso. Com isso, o próximo classificador deverá proceder com uma melhor classificação dessas instâncias que o classificador anterior e assim sucessivamente. Ao final desse processo, os classificadores que possuírem melhor desempenho, ou seja, os que possuem melhor acurácia recebem pesos ( $\alpha_m$ ) maiores em relação aos outros. Assim que todos os classificadores são treinados, eles são combinados na forma de um comitê de classificadores ( $Y_M$ ) (BISHOP, 2006).





**Figura 12- Esquema lógico do algoritmo Adaboost**

Fonte: Bishop (2006).

A lógica por trás do algoritmo do *adaboost* faz com que cada modelo criado dependa do resultado do modelo anterior. Com a ponderação de instâncias é possível fazer com que o modelo foque em determinadas instâncias do conjunto de dados. Assim, cada modelo novo se torna “especialista” em classificar as instâncias que foram marcadas com classificação errada pelo modelo anterior, melhorando, assim, a cada iteração, o modelo proposto (WITTEN; FRANK, 2005).

### 3.4.2. Naive Bayes

*Naive Bayes* é um classificador conhecido por ser simples e eficaz. Esse modelo de classificação é baseado no teorema de Bayes. O nome do algoritmo possui o adjetivo “*Naive*” (ingênuo em inglês) pelo fato de que a hipótese de independência entre os atributos dificilmente é satisfeita. Mesmo com essa hipótese violada, o classificador ainda apresenta bom desempenho (RASCHKA, 2014). Esse algoritmo utiliza as frequências e combinações que ocorrem em um conjunto de dados e as aplica no teorema de Bayes. É comum que os atributos do banco de dados sejam do tipo categóricos para aplicação dessa técnica, porém ela também pode ser usada em dados contínuos desde que esses sejam convertidos em dados categóricos (BRAMER, 2016).

O teorema de Bayes é uma equação usada para calcular a probabilidade condicional ou a posteriori que é a probabilidade de um evento ocorrer dado que uma

variável, ou atributo, possui um determinado valor (BRAMER, 2016). No contexto de problemas de classificação, o teorema de Bayes determina a probabilidade de um objeto pertencer a uma classe quando seus atributos são conhecidos (RASCHKA, 2014). Na forma de equação o teorema toma a forma descrita na equação 1 (ZHANG, 2004):

$$P(c|E) = \frac{P(E|c).P(c)}{P(E)} \quad \text{Eq. 1}$$

Onde:  $P(c|E)$  é a probabilidade de o objeto “E” pertencer a classe “c”.  $P(E|c)$  é a probabilidade da classe c ocorrer dado o objeto E.  $P(c)$  é a probabilidade de ocorrência da classe c e  $P(E)$  é a probabilidade de ocorrência do objeto E dentro do conjunto de dados.

*Naive Bayes* é um algoritmo de classificação cujos resultados são muito bons se comparados a outras técnicas. Mesmo com a premissa de independência entre os atributos não ser satisfeita em grande parte das aplicações, esse método ainda pode apresentar resultados positivos e ser superior a outras técnicas. Uma hipótese para justificar os bons resultados do *Naive Bayes*, mesmo em casos em que haja dependência entre os atributos e classes, é que essa dependência se encontra igualmente distribuída nos atributos e classes presentes. Sendo assim, o que afeta a classificação desse algoritmo seria a combinação de dependências entre os atributos e classes e não o fato de haver dependência em si (ZHANG, 2004).

O código em R para aplicação do classificador *Naive Bayes* possui diversos parâmetros para processar de maneira adequada os dados do conjunto de dados utilizado. As distribuições de classe disponíveis são: Bernoulli, Categórica, Gaussiana, Poisson e *Kernel Density Estimation* (KDE) para representações não-paramétricas. O Bernoulli *Naive Bayes* é aplicado para dados com valores 0 e 1 que sejam modelados por uma distribuição de Bernoulli. O *Gaussian Naive Bayes* funciona em dados que são números reais e são modelados por distribuições gaussianas. Para atributos lógicos, categóricos e com caracteres, a distribuição categórica é usada. Para dados que são números reais e probabilidades condicionais que são estimadas de maneira não-paramétricas, é aplicada a KDE (MAJKA, 2020).

### 3.4.3. Modelos de Misturas Gaussianas (GMM)

Uma distribuição Gaussiana é simétrica em relação à sua média e pode ser definida por sua média e desvio padrão. Muitas aplicações cotidianas possuem dados com distribuições complexas que não se adequam a uma distribuição Gaussiana unimodal. Para descrever dados com distribuições variadas, o GMM utiliza diversas distribuições

gaussianas para representar com mais eficácia os padrões presentes no conjunto de dados observados. Dessa forma, o GMM associa probabilisticamente dados a *clusters* e cada *cluster* é aproximado por uma distribuição Gaussiana distinta, fazendo com que o *cluster* gerado preserve a heterogeneidade dos padrões contido nos dados observados (PATEL; KUSHWAHA, 2020). Assim, mediante a superposição de distribuições Gaussianas, chamadas de componentes da mistura, o GMM pode se adaptar a diversas distribuições, mesmo em dados onde a distribuição dos dados é complexa ou desconhecida (CHEN; WU, 2019).

O algoritmo do GMM utiliza o método de aprendizado não supervisionado para particionar os dados em *clusters*. Considerando um conjunto de dados com N observações ou instâncias, onde  $x_n$  é o vetor de entrada e  $y_n$  é o vetor de saída, uma função de densidade de  $(x_n, y_n)$  que engloba os vetores de entrada e saída, pode ser escrita como uma mistura de k distribuições Gaussianas N como mostra a equação 2. Dessa forma, os elementos probabilísticos do conjunto de dados são divididos em k distribuições Gaussianas, levando em o coeficiente de ponderação, média e covariança de cada distribuição (KARAMI; WANG, 2018).

$$f(x, y | \pi, \mu, \delta) = \sum_{j=1}^k \pi_j N(x, y | \mu_j, \delta_j) \quad \text{Eq. 2}$$

Onde:

$$\sum_{j=1}^k \pi_j = 1, \mu_j = \begin{bmatrix} \mu_{jx} \\ \mu_{jy} \end{bmatrix}, \delta_j = \begin{bmatrix} \delta_{jxx} & \delta_{jxy} \\ \delta_{jyx} & \delta_{jyy} \end{bmatrix} \quad \text{Eq. 3}$$

Devido à facilidade de implementação e simplicidade, o GMM é um método de classificação muito usado em imagens e vídeos, em especial na área de geração e transmissão de imagens e vídeos, na qual são gerados sinais aleatórios que, geralmente, podem ser representados por distribuições gaussianas (XUE; JIANG, 2018). Por sua capacidade de se adaptar a diversas distribuições, também é aplicado para modelagem e monitoramento de dados reais (CHEN; WU, 2019, DIAZ-ROZO; BIELZA; LARRAÑAGA, 2020). Em comparação com outros algoritmos de agrupamento, como o K-Means, o GMM dispõe mais informações de cunho quantitativo e apresenta mais opções de interpretação acerca do comportamento e dos padrões dos dados observados (DIAZ-ROZO; BIELZA; LARRAÑAGA, 2020).

#### 3.4.4. Árvore de decisão

A árvore de decisão é um algoritmo comumente aplicado na área de inteligência artificial e aprendizado de máquina pois promove uma representação simples de

informações adquiridas a partir dos dados observados, o que possibilita a compreensão de diversos públicos, não apenas de especialistas (TRABELSI; ELOUEDI; LEFEVRE, 2018). Além disso, a árvore de decisão é um algoritmo de fácil implementação e robusto, ou seja, pode ser usado em dados categóricos e numéricos e pode ser usada em bancos de dados extensos. O principal objetivo de uma árvore de decisão é criar um modelo capaz de prever a classe ou o valor de um atributo alvo para instâncias ainda não conhecidas, baseando-se em atributos conhecidos de entrada (FARID *et al*, 2014).

O algoritmo de construção de uma árvore de decisão consiste na estratégia “dividir para conquistar”. Para um conjunto de dados com diversos atributos, o algoritmo escolhe o atributo que, de acordo com métricas como: ganho de informação e razão de ganho, permite o melhor particionamento dos dados de treino. A partir desse atributo, são separados um conjunto de dados de treino e subconjuntos de treino (TRABELSI; ELOUEDI; LEFEVRE, 2018). Considerando um conjunto de dados hipotéticos, contendo três atributos e duas classes chaves, classe X e Y, que poderiam corresponder a classificações como “positivo” e “negativo”; “aberto” e “fechado”; “falha” e “normal”, observa-se na figura 13, que o atributo escolhido para partição dos dados foi o atributo 1. Esse atributo possui três instâncias. Cada segmento constitui uma regra a ser seguida; se o atributo 1 corresponder à opção A é necessário utilizar o atributo 2, outro ramo da árvore, para se obter a classificação alvo dessa instância; se corresponder à opção B a classe só pode ser X, sendo assim é alcançada a folha da árvore onde não é necessário nenhuma informação adicional para a classificação da instância. Caso o atributo 1 seja correspondente à opção C, então é necessário utilizar mais uma ramificação, o atributo 3, para determinar a classificação da instância.

A Figura 13 apresenta um exemplo genérico de árvore de decisão na qual os atributos são do tipo categórico. No entanto, uma árvore de decisão também pode ser aplicada em atributos numéricos. Nesse caso, cada ramificação da árvore apresenta a comparação entre o valor do atributo e uma constante, é possível também que dois ou mais atributos sejam comparados entre si. As folhas das ramificações apresentam a classificação que é atribuída a todas as instâncias que percorrem a árvore através das regras de comparação impostas em cada ramificação e chegam àquela folha (WITTEN; FRANK, 2005).

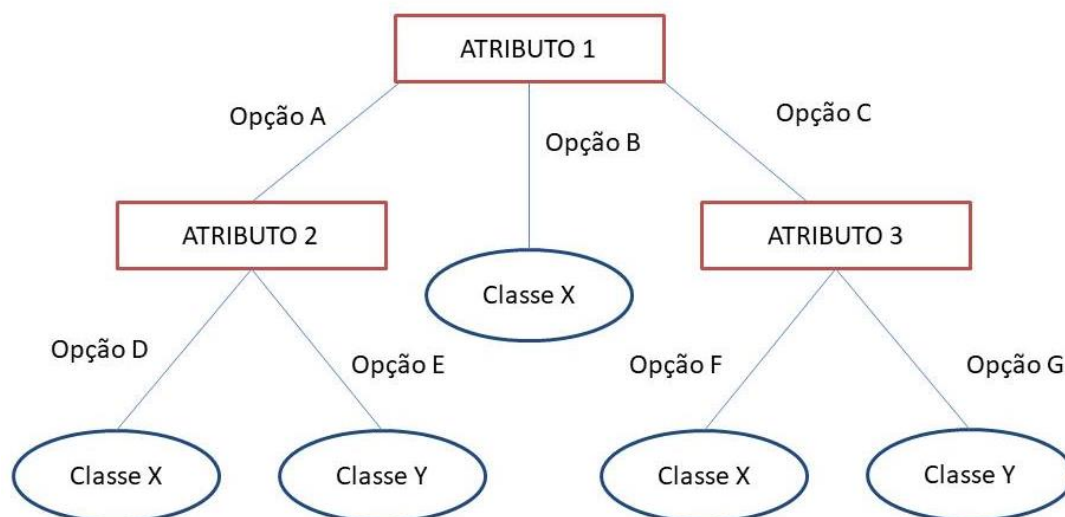


Figura 13- Exemplo de árvore de decisão genérica

Fonte: a Autora.

As árvores de decisão podem sofrer com a baixa qualidade do banco de dados usado, instabilidade e problemas de generalização. Sendo assim, é fundamental processar os dados para obter melhores resultados (FARID *et al*, 2014).

Há diversos algoritmos disponíveis para construção de árvores de decisão. Os tipos clássicos de árvores de decisão são: CART, C4.5, C5.0 e NBTree. A árvore CART (*Classification And Regression Tree*) é um método que comporta classificação e regressão, ela constrói uma árvore binária e divide os dados de treino de acordo com o índice gini (*gini index*). A partir disso são construídas árvores de regressão que predizem a classe de uma dada instância. O algoritmo C4.5 seleciona os atributos com base na razão de ganho de informação e evita escolhas tendenciosas (*bias*) durante o processo de seleção. Ele pode ser usado em atributos discretos e contínuos, e consegue obter bons resultados mesmo em bancos com dados faltantes. A versão mais avançada do algoritmo C4.5 é o C5.0, que agrega aspectos como o sistema de voto *boosting* na seleção dos atributos. Ele também é mais eficiente na geração de árvores, pois descarta árvores que não contribuem para a melhoria da acurácia de classificação. Já o NBTree (*Naive Bayes Tree*) é um algoritmo híbrido que combina uma árvore de decisão clássica com o classificador *naive Bayes* nas folhas ao invés da determinação de uma única classe (GUGGARI; KADAPPA; UMADEVI, 2018)

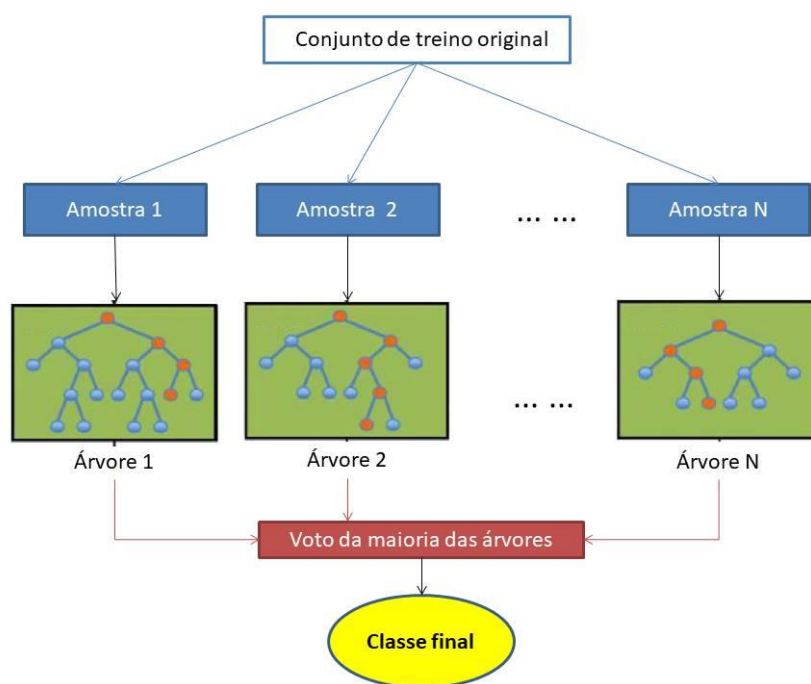
### 3.4.5. *Random forest*

Construído com base na árvore de decisão CART, o algoritmo *random forest* agrega diversas árvores de decisão e, por isso, apresenta melhores resultados que uma única árvore. Pode ser usado em problemas de classificação e regressão (YOO *et al*, 2019). Além disso, o *random forest* possui como vantagens ser de fácil visualização e possuir baixo custo computacional (LAZRI; AMEUR, 2018).

O algoritmo *random forest* cria diversas árvores do tipo CART e utiliza dois métodos para reduzir a variância e controlar comportamentos tendenciosos (LIU; GE, 2018). O primeiro é chamado de *bagging* que engloba o processo de *bootstrap*. Esse método consiste em colher amostras aleatórias do conjunto de dados originais, gerando diversos subconjuntos distintos entre si (MARTLOF, 2017). Em seguida, além da escolha aleatória do subconjunto de dados a ser usado no treino das árvores de decisão que irão compor a “floresta”, os atributos a serem testados nos ramos das árvores também são escolhidos de forma aleatória (LIU; GE, 2018), diferente do ocorre numa árvore de decisão tradicional.

A Figura 14 detalha os procedimentos adotados durante a execução do algoritmo *random forest*. Primeiramente, a partir do conjunto de treino original, são criadas N amostras aleatórias distintas pelo método de *bootstrap*. Então, para cada subconjunto de treino é criada uma árvore do tipo CART, onde a seleção dos atributos nos ramos das árvores também é feita de forma aleatória. Depois de diversas árvores serem geradas e treinadas, as instâncias a serem classificadas são preditas com a contabilização dos votos das árvores, classe votada pela maioria determina a classificação final da instância. Em problemas de regressão, ao invés do sistema de votação, é levada em conta a média dos valores obtidos pelas árvores para determinar o valor da instância (LIU; GE, 2018).

A técnica de *random forest* é efetiva em encontrar relações entre o atributo a ser explicado e os atributos explicativos. Por meio dele é possível classificar os atributos explicativos a partir de sua conexão com a variável explicada. Há apenas dois parâmetros a serem estabelecidos para aplicação do algoritmo; o número de árvores a serem criadas na “floresta” e o número de variáveis nos ramos das árvores em cada subconjunto (LAZRI; AMEUR, 2018). Esses dois parâmetros podem ser otimizados para criar florestas de tamanho ideal, sem perder robustez e eficiência computacional. Além disso, o algoritmo possui boa capacidade de generalização e não apresenta problemas de *overfitting* (GHASEMIAN; AKHOONDZADEH, 2018).



**Figura 14- Algoritmo Random Forest**

Fonte: adaptado de Liu e Ge (2018).

Assim como algoritmo *adaboost*, o *random forest* também usa o sistema de votos para determinar a classe final de uma instância. No entanto, este inclui elementos de aleatoriedade que não estão presente no *adaboost*, por isso pode ser mais efetivo e apresentar melhor acurácia que o *adaboost* (SHIPWAY *et al*, 2019).

### 3.5. Métricas de avaliação

Além dos algoritmos citados anteriormente, ainda existem diversas outras técnicas de classificação. Algumas demandam mais tempo de processamento, exigem mais memória ou necessitam de conjuntos de treinos maiores para produzir resultados confiáveis. Para avaliar cada algoritmo e conhecer qual é o melhor a ser aplicado podem ser utilizados métodos para determinação da acurácia de predição (BRAMER, 2016).

Com esse propósito deve ser elaborada uma matriz de confusão, que mostra quantas vezes as instâncias de uma determinada classe foram classificadas corretamente e quantas vezes o algoritmo errou a classificação, detalhando o desempenho do algoritmo. Na matriz de confusão estão dispostas informações obtidas pelo conjunto de treino e de teste. A tabela 1 apresenta uma matriz de confusão genérica que mostra a classe predita pelo algoritmo a partir do conjunto de treino e a classe verdadeira, obtida pelas classes observadas no conjunto de teste para validação dos resultados (BRAMER, 2016).

No exemplo proposto na Tabela 1 existem apenas duas classes positivo (+) e negativo (-), mas o exemplo pode ser extendido para um número maior de classes. Na diagonal principal encontram-se as instâncias que foram classificadas corretamente, os verdadeiros positivos (VP) e verdadeiros negativos (VN). Na diagonal secundária, estão as instâncias em que o algoritmo não acertou a classificação. Instâncias cuja classe era positivo e foi classificada pelo algoritmo como negativo, recebem o nome de falso negativo (FN). Por outro lado, instâncias cuja classe verdadeira era negativo e o algoritmo classificou como positivo são chamadas de falsos positivos (FP). Na última coluna da tabela 1 encontra-se o total de instâncias de cada classe, calculado pela soma das linhas da matriz.

**Tabela 1- Matriz de confusão**

		Classe predita pelo algoritmo		Total de instâncias
		+	-	
Classe verdadeira	+	VP	FN	Nº de instâncias “positivo” (P)
	-	FP	VN	Nº de instâncias “negativo” (N)

Fonte: Adaptado de Bramer (2016).

Falsos positivos e falsos negativos geram dois tipos de erro. O erro do tipo 1 está relacionado à quantidade de falsos positivos, ou seja, quando o dado é da classe “negativo” e foi classificado como positivo. Já o erro do tipo 2 está ligado aos falsos negativos, dados que são da classe “positivo” e foram classificados como negativo. Dependendo da aplicação, é fundamental conhecer o tipo de erro mais frequente cometido pelo classificador, pois cada tipo de erro possui impactos distintos. Para uma aplicação pode ser de interesse minimizar o número de falsos positivos para obter melhores resultados enquanto para um outro tipo de aplicação seja mais vantajoso minimizar a quantidade de falsos negativos (BRAMER, 2016).

A avaliação de desempenho de uma técnica de classificação é dada por métricas que consideram a quantidade de falsos positivos (FP) e negativos (FN), bem como o número total de instâncias de cada classe (P e N). As principais métricas usadas em diversas áreas são descritas a seguir (BRAMER, 2016):

- *Recall*: também conhecido como taxa de verdadeiros positivos ou *sensitivity*, é calculado pela divisão  $VP/P$  e representa a proporção de instâncias positivas classificadas de forma correta.



- Taxa de falsos positivos: conhecido como falso alarme, é dado pela razão  $FP/N$ . Considera a proporção de instâncias “negativo” classificadas como “positivo”.
- Taxa de falsos negativos: é dado pela razão  $FN/P$ , complementar à taxa de falsos positivos.
- Taxa de verdadeiros negativos: também recebe o nome de *specificity*, é calculado pela razão  $VN/N$  e representa a proporção de instâncias de classe “negativo” que foram classificados como tal.
- *Precision*: representa a proporção de instâncias classificadas como “positivo” que são realmente “positivo”, é dada pela fórmula  $VP/(VP+FP)$ .
- *F1-score*: engloba *recall* e *precision* e é calculado pela média harmônica  $2 \times \frac{Precision \times Recall}{Precision + Recall}$ .
- Acurácia: representa a proporção de instâncias que foram corretamente classificadas. É calculada pela fórmula:  $(VP+VN)/(P+N)$ .

A acurácia é a métrica mais usada para avaliar classificadores, porém apresenta algumas limitações. Para conjuntos de dados com classes em quantidades desbalanceadas, a acurácia pode mascarar o desempenho do algoritmo. Por outro lado, o *F1-score* é uma boa alternativa de métrica, em problemas de classificação binária pode, inclusive, apresentar melhores resultados que a acurácia (HOSSIN; SULAIMAN, 2015).

# Capítulo 4

## Metodologia

A metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) propõe o desenvolvimento de um projeto de mineração de dados de forma estruturada em seis processos ou etapas (PINTO, 2018). Como parte do trabalho proposto envolve trabalhar com dados monitorados, essa metodologia foi implementada para dar mais clareza aos processos realizados para obtenção dos resultados dessa pesquisa.

A primeira etapa da metodologia CRISP-DM consiste em entender o negócio ao qual o projeto está inserido. No presente estudo, o “negócio” é a produção de petróleo. As variáveis envolvidas são de sensores presentes ao longo do sistema de produção e que permitem o monitoramento da produção. A segunda etapa contempla a compreensão dos dados a serem usados. Os dados de produção utilizados nesse trabalho são oriundos de um banco de dados publicado por Vargas et al (2019), onde os autores disponibilizaram dados reais, simulados e “desenhados” a mão. Nesse banco de dados os autores colocaram horas de monitoramento de sensores de pressão temperatura e vazão de *gas lift* durante a ocorrência de anomalias indesejadas na produção de petróleo. O banco de dados é dividido de acordo com a anomalia em questão. Cada anomalia apresenta um conjunto de dados reais ou simulados em arquivos distintos, com datas de ocorrência diferentes. A terceira etapa é a de preparação dos dados. Nessa etapa foram selecionados os poços com dados reais de hidratos, incrustação, *slugging* e instabilidade de fluxo que apresentavam maior qualidade de dados, ou seja, foram escolhidos aqueles com mais informações de sensores disponíveis. A etapa seguinte é a de modelagem, onde foram escolhidos os pacotes e abordagens a serem utilizados na construção dos modelos de classificação. Nesta fase, os modelos foram treinados e testados nos experimentos dos blocos 1 e 2. A quinta etapa, designada como “avaliação”, analisa os modelos gerados e avalia seus desempenhos. Essa etapa consiste da discussão dos resultados das aplicações dos modelos em cada um dos experimentos propostos. Aqui devem ser avaliados o desempenho dos algoritmos e a forma como os mesmos erraram ao dividir as classes do conjunto de teste, percebendo quais classes eram mais complicadas de serem classificadas. Por último, a sexta etapa, é a implementação dos modelos desenvolvidos no negócio. Como essa pesquisa não desenvolveu um protótipo que pudesse ser acoplado a um sistema de

monitoramento real, não foi possível testar os modelos em condições reais de produção de petróleo.

Discorrida as etapas realizadas com base na metodologia CRISP-DM, agora será abordado com mais detalhes a metodologia de pesquisa, as variáveis e os experimentos realizados de modo a responder as questões propostas no primeiro capítulo.

Nesta dissertação será usado um banco de dados público que apresenta dados reais e simulados, contendo sete tipos de problemas indesejados durante a produção de petróleo. Os problemas relatados são: aumento abrupto do BSW (*Basic Sediments and Water*), fechamento esporádico da válvula DHSV (*Downhole Safety Valve*), *slugging* severo, instabilidade de fluxo, perda rápida de produtividade, restrição rápida na válvula de produção, incrustação na válvula de produção e hidrato na linha de produção (VARGAS *et al*, 2019). Dentre os problemas listados, serão abordadas as anomalias relacionadas à garantia de escoamento. Sendo assim, somente serão usados os dados reais de *slugging* severo, instabilidade de fluxo, incrustação na válvula de produção e formação de hidratos na linha de produção.

O banco de dados é baseado em dez variáveis, como mostra a Tabela 2, que apresenta as variáveis relacionadas aos sensores de pressão e temperatura. Além desses sensores, têm-se as variáveis tempo e classe. A variável classe representa três regimes; normal, transiente e anormal. A classe normal é representada por “0”, o transiente é representado “10X” onde X é o número da anomalia, por último a classe que representa o comportamento anormal é representada pelo “X” onde X varia de 1 a 8, pois existem 8 problemas distintos. No banco de dados, os problemas 3, 4, 7 e 8 referem-se a *slugging*, instabilidade de de fluxo, incrustação e hidrato, respectivamente.

**Tabela 2- Identificação de sensores e variáveis relacionadas**

Sigla	Nome	Especificação
PDG	<i>Permanent Downhole Gauge</i>	Sensor de pressão localizado na no interior do poço
P-TPT	<i>Pressure at the Temperature and Pressure Transducer</i>	Sensor de pressão localizado na árvore de natal
T-TPT	<i>Temperature at the Temperature and Pressure Transducer</i>	Sensor de temperatura localizado na árvore de natal

P-MON-CKP	<i>Pressure upstream the choke production valve</i>	Sensor de pressão à montante da válvula <i>choke</i> de produção, localizado no <i>topside</i> da UEP
T-JUS-CKP	<i>Temperature downstream the choke production valve</i>	Sensor de temperatura à jusante da válvula <i>choke</i> de produção, localizado no <i>topside</i> da UEP
P-JUS-CKGL	<i>Pressure downstream choke gas lift valve</i>	Sensor de pressão à jusante da válvula <i>choke</i> de <i>gas lift</i> , localizado no <i>topside</i> da UEP
T-JUS-CKGL	<i>Temperature downstream choke gas lift valve</i>	Sensor de temperatura à jusante da válvula <i>choke</i> de <i>gas lift</i> , localizado no <i>topside</i> da UEP
QGL	<i>Gas lift flow</i>	Vazão de injeção de <i>gas lift</i>

O banco de dados apresenta as oito anomalias em pasta distintas, cada pasta contém arquivos reais e simulados de alguns poços produtores. Os dados reais de *slugging*, instabilidade de fluxo e incrustação não possuem a variável T-JUS-CKGL. Os dados de hidrato não possuem a variável T-JUS-CKP e nem T-JUS-CKGL. Portanto, essas variáveis serão excluídas no processo de pré-tratamento dos dados, assim como instâncias vazias também serão retiradas.

Pela questão da similaridade física, os quatro problemas serão divididos em dois blocos e diversos experimentos, como mostra a Tabela 03. Como tanto hidrato quanto incrustação geram redução do diâmetro disponível para escoamento, porém em posições distintas, eles formarão um bloco de análise. Já o outro bloco de análise aborda a instabilidade de fluxo e o *slugging* que são problemas correlatos, instabilidade de fluxo é caracterizada por mudanças significativas em pelo menos uma das variáveis monitoradas (GUERRERO-SARABIA; FAIRUZOV, 2013), mas não possui a característica de periodicidade do *slugging*. Essa separação permite uma análise mais detalhista pois considera que, durante o monitoramento da produção, a distinção desses problemas poderia ser difícil.

Tabela 3– Experimentos da dissertação

<b>Bloco 1: hidrato/incrustação</b>	<b>Bloco 2: instabilidade de fluxo/<i>slugging</i></b>
<b>Experimento 1:</b> aplicar as técnicas de aprendizado de máquina para detectar hidratos utilizando um conjunto de treino e teste proveniente do mesmo poço.	<b>Experimento 1:</b> aplicar série temporal nos dados de instabilidade de fluxo e <i>slugging</i> de um poço e comparar com a condição normal antes e depois da ocorrência das anomalias.
<b>Experimento 2:</b> aplicar as técnicas de aprendizado de máquina para detectar hidratos utilizando os modelos do experimento 1 e classificar um poço diferente do usado no treinamento (diferente conjunto de teste).	<b>Experimento 2:</b> aplicar as técnicas de aprendizado de máquina em uma base de dados mesclada, com dados de instabilidade de fluxo e <i>slugging</i> , usando a mesma como treino e teste.
<b>Experimento 3:</b> aplicar as técnicas de aprendizado de máquina para detectar incrustação utilizando um conjunto de treino e teste proveniente do mesmo poço.	<b>Experimento 3:</b> aplicar as técnicas de aprendizado de máquina em uma base de dados mesclada, com dados de instabilidade de fluxo e <i>slugging</i> , e testar com dados de instabilidade de fluxo e <i>slugging</i> de um poço diferente do utilizado no treinamento
<b>Experimento 4:</b> aplicar as técnicas de aprendizado de máquina para detectar incrustação utilizando os modelos do experimento 3 classificar um poço diferente do usado no treinamento (diferente conjunto de teste).	

Para bloco de análise de hidrato e incrustação serão realizados quatro experimentos. Os primeiros dois experimentos, referentes aos poços que contém hidratos, são de aplicação dos algoritmos de classificação com os dados de treinamento e teste de um mesmo poço e, posteriormente, a aplicação de dados de um poço diferente do usado no treinamento como conjunto de teste, totalizando dois experimentos. A segunda fase de experimentos, referentes aos poços que apresentam incrustação na válvula *choke* de

produção, replicam as ações realizadas nos experimentos anteriores, também totalizando dois experimentos.

Para o bloco de análise de instabilidade de fluxo e *slugging*, primeiramente será realizada uma análise temporal para determinar quais sensores captam o comportamento periódico do *slug* e também verificar se a instabilidade de fluxo apresenta algum padrão temporal. Após a análise temporal, serão aplicados os algoritmos de classificação num banco de dados mesclado artificialmente com esses dois problemas. Nem instabilidade de fluxo, nem *slugging* apresentam um arquivo contendo as condições normais e transientes juntas. Isso se acontece pelo fato de não haver estado transiente para esses problemas, ou o fluxo está normal ou instável. Sendo assim não será possível realizar a classificação dos estados normal, transiente e anormal. Por isso, serão mesclados dados de instabilidade de fluxo e *slugging* para treinar os algoritmos de classificação. Para este bloco serão realizados três experimentos. Além da aplicação de séries temporais, os dois outros experimentos utilizarão a base de dados mesclada. O segundo experimento utilizará um poço para treinar e testar os modelos de classificação e, posteriormente, no último experimento, um poço distinto será usado para testar os modelos gerado.

A métrica de análise do desempenho dos algoritmos será realizada através da matriz de confusão e do *F1-score* de cada experimento. Neste caso, a escolha do *F1-score* em detrimento da validação cruzada (*cross validation*), muito usada para avaliação de modelos, se dá pelo fato dos eventos de interesses representarem uma parcela muito pequena do banco de dados, pois são eventos considerados raros. Sendo assim, há a possibilidade da validação cruzada não utilizar um conjunto de teste que contenha a classe de interesse.

No bloco hidrato/incrustação, será verificada a eficiência de cada algoritmo em classificar os regimes normal, transiente e anormal de um poço e sua viabilidade em classificar poços diferentes. Para o bloco instabilidade de fluxo/*slugging* será avaliada a característica temporal de cada um dos sensores, também será usada uma base de treino e teste contendo dados de ambos os problemas. Posteriormente, serão usados dados de poços diferentes contendo essas duas anomalias. O intuito é assegurar que os algoritmos detectem dados de *slugging* ou instabilidade de fluxo nos dados de produção de um poço e verificar se o modelo construído a partir de um poço pode ser estendido para outros poços.

A principal razão para separação em blocos dos problemas, além da natureza dos problemas, é a disponibilidade de amostras do banco de dados. Não há poços que

apresentem tanto hidrato quanto incrustação. Sendo assim, a distinção entre essas duas anomalias também será feita através dos resultados obtidos pela aplicação do *naive Bayes* que mostra a distribuição dos valores e os valores médios das variáveis em cada uma das classes estudadas.

Por outro lado, para instabilidade de fluxo e *slugging*, que são eventos correlatos, existem poços que apresentam ambos os problemas. Dessa forma, é possível agrupá-los e fazer uma base de dados mista de treino. Assim, as ferramentas de classificação poderão ser usadas para realizar a distinção entre essas anomalias num mesmo conjunto de dados, como será observada na matriz de confusão e *F1-score* dos algoritmos ao identificar cada anomalia usando a base de treino mista.

#### 4.1. Metodologia experimental bloco 1

Todos os experimentos apresentados neste trabalho foram desenvolvidos na plataforma gratuita “Rstudio”. O *script* utilizado no bloco 1 possui mais etapas de limpezas, uma vez que os poços contidos no banco de dados, que apresentam formação de hidratos e incrustação possuem mais variáveis. Nos poços de hidratos, sete das dez variáveis são realmente utilizadas. Enquanto nos poços que apresentam incrustação, seis variáveis são usadas nos modelos.

A primeira etapa na realização dos experimentos 1 e 2 deste bloco é a instalação dos pacotes necessários. A tabela 4 explicita todos os pacotes utilizados em ambos os blocos de análise. Em seguida, é realizada a limpeza dos dados. O objetivo é eliminar instâncias vazias e variáveis nulas.

Posteriormente, são criados aleatoriamente dois conjuntos de dados, dados de treino e teste. Os dados de treino correspondem a 70% do conjunto original, enquanto os dados de teste equivalem aos 30% restante. Os modelos, criados a partir das funções disponíveis nos pacotes instalados, são treinados com o conjunto de treino e testados com dados diferentes dos usados no treinamento, o conjunto de teste, sem o atributo “classe”. Então, a previsão dada pelo modelo é comparada com a classe verdadeira contida na coluna “classe” dos dados de teste. Com isso, por meio da matriz de confusão, é comparada a previsão do modelo com a referência determinada pelo atributo “classe”. A partir da matriz de confusão é possível obter a acurácia e o *F1-score* de cada modelo e analisar a sensibilidade do modelo quanto à identificação da saída do padrão normal para o transiente e saída do transiente para a situação de falha, de anormalidade em si.

Nos experimentos 2 e 4, para avaliar a eficiência dos modelos com dados de poços distintos dos usados no treinamento, serão usados poços que apresentem a mesma anomalia e número de variáveis que o conjunto de treino. Todavia, por se tratarem de poços distintos, foi realizada uma normalização dos dados dos poços usados nessa comparação. O processo de normalização consiste em dividir todas as instancias do conjunto de dados de um determinado poço pela média dos valores da classe normal. Com isso, a condição normal tem seus valores sempre próximo de 1, enquanto as outras classes apresentam valores maiores ou menores que 1. Em relação a métrica de desempenho, a análise também será baseada na matriz de confusão e no *F1-score*.

**Tabela 4- Pacotes em R necessários para a realização dos experimentos**

<b>Pacote</b>	<b>Descrição</b>
caret	Funções de treinamento e ferramentas gráficas para modelos de classificação e regressão (CRAN, 2020).
naivebayes	Implementação do classificador Naive Bayes. Contém as seguintes distribuições condicionais de classe: Bernoulli, Categórica, Gaussiana, Poisson e representação não paramétrica de densidade condicional de classe estimada via <i>Kernel Density Estimation</i> (CRAN, 2019).
adabag	Implementa Freud e Schapire Adaboost. Algoritmos M1 e <i>Breiman's Bagging</i> são usados, utilizando árvores classificadoras como classificadores individuais (CRAN, 2018).
janitor	Formata o banco de dados. Organiza o nome das colunas, possibilita contagem rápida de diversos atributos e isola dados duplicados. Outra função do pacote é formatar a tabulação dos resultados do modelo (CRAN, 2020).
bnlearn	Implementa <i>constraint-based</i> , <i>pairwise</i> , <i>score-based</i> e estruturas híbridas de algoritmos de aprendizado estruturado para redes discretas, gaussianas e condicionais gaussianas, além de outras funções e testes (CRAN, 2019).
rpart.plot	Extende as funcionalidades do pacote “rpart”, este possui recursos de partição de dados para classificação e árvores de regressão (CRAN, 2019).



randomForest	Realiza classificação e regressão baseadas em florestas de árvores, usando <i>inputs</i> aleatórios. (CRAN, 2018).
ggplot2	Promove a criação de gráficos a partir da estrutura escolhida pelo usuário (CRAN, 2019).
reshape2	Reestrutura e agrega dados de maneira flexível (CRAN, 2017).
corrplot	Representa graficamente matrizes comuns ou de correlação. Possui também artifícios para reorganização de matrizes e edição gráfica (CRAN, 2017).

## 4.2. Metodologia experimental do bloco 2

Assim como no bloco 1, os pacotes utilizados nesta análise também se encontram explicitados na tabela 4. As etapas de limpeza de dados e a proporção usada na separação do conjunto de treino e teste também são semelhantes.

O experimento 1 deste bloco refere-se á análise temporal do *slugging*, da instabilidade de fluxo e das condições normal antes e depois das anomalias. A análise temporal é reponsável por datalhar o comportamento dos sensores do poço ao longo do tempo, mostrando se há ou não algum padrão. Nesta etapa não há separação de conjunto de treino e teste, todos os dados do poço são analisados cronologicamente.

No experimento 2, utiliza-se uma base de dados misturada manualmente, com dados de instabilidade de fluxo e *slugging*, oriunda de um único poço que apresentou esses dois problemas de garantia de escoamento em dias diferentes. Essa base de dados mista será utilizada para treinar e testar os modelos e, através do F1-score e da matriz de confusão será feita a avaliação do desempenho dos modelos nesta base de dados. Já o experimento 3 utiliza os modelos de classificação criados no experimento 2 para classificar os dados de *slugging* e instabilidade de fluxo de um poço diferente do usado no treinamento do experimento anterior.

# Capítulo 5

## Resultados e discussão

Este capítulo apresenta os resultados obtidos a partir dos experimentos realizados e a discussão desses resultados. O capítulo será dividido de acordo com o bloco de experimentos. Primeiramente serão abordados os resultados do bloco 1 e algumas análises necessárias para gerar os modelos de classificação. Em seguida, será discutido o segundo bloco de experimentos.

### 5.1. Bloco de experimentos formação de hidrato e incrustação

Antes de aplicar os algoritmos de aprendizado de máquina é importante conhecer os dados que serão utilizados. Neste estudo são utilizados dados de sensores localizados dentro do poço, no leito marinho e no *topside* de uma UEP que fornecem informações sobre a produção de petróleo de um poço. Para problemas de formação de hidrato, dentre os três poços disponíveis, serão utilizados dois poços (chamados de 20 e 21 no banco de dados) pois esses apresentam maior número de variáveis e melhor qualidade de dados. Por outro lado, para incrustação dentre os três diferentes poços disponíveis apenas um poço continha as três classes usadas na análise. Outro poço apresentava apenas as classes normal e transiente e, no outro apenas o transiente. Sendo assim, no estudo comparativo foram usados o poço com todas as classes (01) e o poço com classe normal e transiente (06).

#### 5.1.1. Análise dos poços

Anteriormente à aplicação das técnicas de classificação realizou-se um estudo comparativo entre os dois poços a serem utilizados, com o intuito de verificar como a formação de hidrato progride em cada poço. Ambos os poços estudados não possuem as variáveis T.JUS.CKGL e T.JUS.CKP, sendo assim essas variáveis não são apresentadas nos resultados. A data de ocorrência de hidrato no poço 20 é de abril de 2012, enquanto no poço 21 é de maio de 2017. Outro ponto de análise está no fato do poço 20 possuir uma janela de dados disponível menor que o poço 21.

A partir do sumário do poço 20 e 21 mostrados nas Figuras 15 e 16, respectivamente, é possível a amplitude dos valores de cada variável avaliada em cada um dos poços. O patamar de pressão nos sensores *subsea* do poço 21 é maior que o poço 20, o que pode ter origem em diversas causas, afinal poços diferentes podem ser oriundos

de reservatórios diferentes, possuírem configurações geométricas distintas e não apresentar o mesmo tempo de produção, por exemplo.

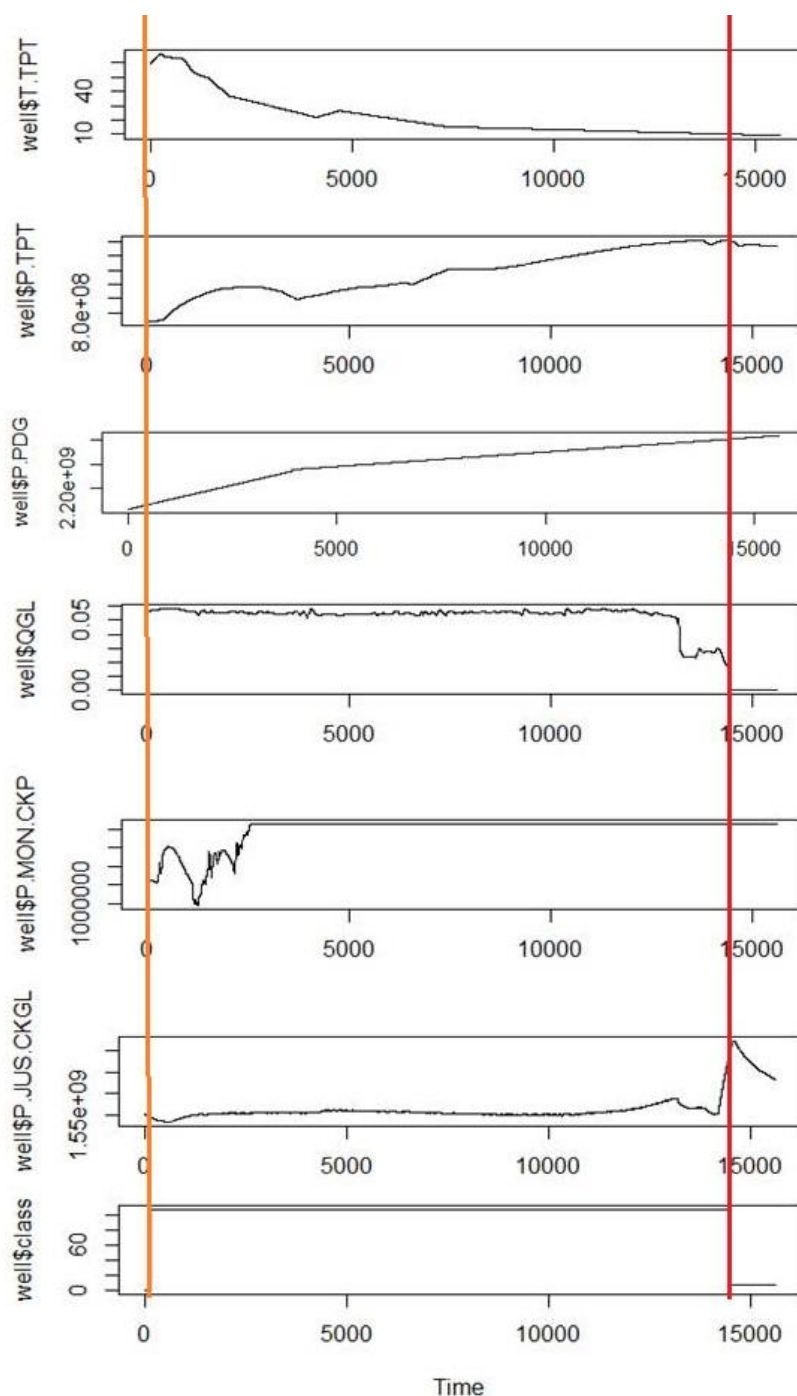
timestamp	P.PDG	P.TPT	T.TPT
Length:15602	Min. :2.157e+09	Min. :6.739e+08	Min. : 9.25
Class :character	1st Qu.:2.235e+09	1st Qu.:1.138e+09	1st Qu.:12.11
Mode :character	Median :2.263e+09	Median :1.405e+09	Median :14.96
	Mean :2.255e+09	Mean :1.380e+09	Mean :21.92
	3rd Qu.:2.286e+09	3rd Qu.:1.710e+09	3rd Qu.:25.64
	Max. :2.310e+09	Max. :1.826e+09	Max. :66.42
P.MON.CKP	P.JUS.CKGL	QGL	class
Min. : 942928	Min. :1.532e+09	Min. :0.0002015	Min. : 0.00
1st Qu.:3148975	1st Qu.:1.551e+09	1st Qu.:0.0540183	1st Qu.:108.00
Median :3149025	Median :1.555e+09	Median :0.0550757	Median :108.00
Mean :2970301	Mean :1.565e+09	Mean :0.0489350	Mean : 99.69
3rd Qu.:3149045	3rd Qu.:1.560e+09	3rd Qu.:0.0560572	3rd Qu.:108.00
Max. :3149214	Max. :1.724e+09	Max. :0.0586095	Max. :108.00

Figura 15-Estatística descritiva do poço 20

timestamp	P.PDG
Length:24059	Min. :3.077e+09
Class :character	1st Qu.:3.079e+09
Mode :character	Median :3.085e+09
	Mean :3.085e+09
	3rd Qu.:3.092e+09
	Max. :3.101e+09
P.TPT	T.TPT
Min. :8.496e+08	Min. : 4.311
1st Qu.:1.013e+09	1st Qu.: 7.647
Median :1.618e+09	Median :16.089
Mean :1.448e+09	Mean :13.105
3rd Qu.:1.831e+09	3rd Qu.:17.821
Max. :1.978e+09	Max. :23.343
P.MON.CKP	P.JUS.CKGL
Min. : 897947	Min. :1.364e+09
1st Qu.:1180230	1st Qu.:1.399e+09
Median :1330996	Median :1.465e+09
Mean :1412231	Mean :1.486e+09
3rd Qu.:1552697	3rd Qu.:1.593e+09
Max. :3150695	Max. :1.631e+09
QGL	class
Min. :0.00000	Min. : 0.00
1st Qu.:0.04535	1st Qu.:108.00
Median :0.05394	Median :108.00
Mean :0.05199	Mean : 90.28
3rd Qu.:0.06024	3rd Qu.:108.00
Max. :0.06805	Max. :108.00

Figura 16- Estatística descritiva do poço 21

Analisando temporalmente as variáveis de produção do poço 20 apresentadas na Figura 17, é possível identificar quando o regime começa a entrar no estado de transiência. A partir da linha laranja começa o estado transiente da formação de hidrato que se instala definitivamente após a marcação da linha vermelha. Para esse o poço a condição normal não ficou visível no gráfico porquê esta representa uma proporção muito pequena do banco de dados, sendo suprimida na amostra gráfica.



**Figura 17- Comparação temporal das variáveis de produção do poço 20. A linha laranja indica o início do transiente e a linha vermelha indica a instalação da falha.**

A formação de hidratos ocorre em condições de alta pressão e baixa temperatura. As linhas de produção de petróleo são pressurizadas e o perfil de temperatura vai diminuindo ao longo do escoamento, caso não sejam utilizados isolantes eficazes. A redução da temperatura em poços marítimos se dá principalmente pela perda de calor para ambiente marinho, que se encontra a baixas temperaturas. Observando o comportamento

dos dados do poço 20, verifica-se que ao sair da condição normal de operação a temperatura na árvore de natal cai continuamente atingindo valores inferiores a 10°C; a pressão no fundo do poço e na árvore de natal aumentam mais gradativamente em consequência da contínua redução do diâmetro disponível para escoamento. A pressão à montante da válvula de produção apresenta comportamento difuso no início do transiente e depois aumenta e permanece num patamar mais elevado até a instalação completa da anomalia. Enquanto a pressão à jusante da válvula de *gas lift* apresenta pouca variação durante o transiente e perto da instalação completa da falha, apresenta um aumento brusco. A vazão de *gas lift* apresenta comportamento semelhante, porém, conforme a formação de hidrato se completa, há uma diminuição na vazão de *gas lift*.

Realizando a mesma análise temporal para o poço 21, com base na Figura 18, percebe-se que este poço contém mais dados sobre a condição normal de produção (região antes da linha laranja) e menos dados sobre a condição onde o hidrato está completamente instalado.

A amplitude de valores de temperatura do poço 21 é menor do que a registrada no poço 20. Mesmo assim o comportamento dessa variável durante estado de transiência é similar, apresentando uma queda contínua da temperatura na árvore de natal até atingir valores próximos ao do fundo do mar. A pressão no fundo do poço e na árvore de natal aumentam gradativamente. Diferentemente do que ocorreu no poço 20, a pressão à montante da válvula de produção apresenta comportamento errático ao longo de todo o período transiente e aumenta com a formação completa de hidrato. A pressão à jusante da válvula de *gas lift* apresenta um aumento gradual e a vazão de *gas lift* apresenta uma leve queda durante o transiente e cai ainda mais com a instalação completa da falha.

Comparando os resultados das séries temporais das variáveis de produção dos poços 20 e 21 percebe-se que a pressão no fundo do poço, a temperatura e pressão na árvore de natal, e a vazão de *gas lift* apresentam comportamentos similares em ambos os poços. Já a pressão à montante da válvula de produção e a pressão à jusante da válvula de *gas lift* comportam-se de forma distinta durante o transiente de cada poço, porém o efeito causado pela formação de hidrato é o mesmo, ambos aumentam.

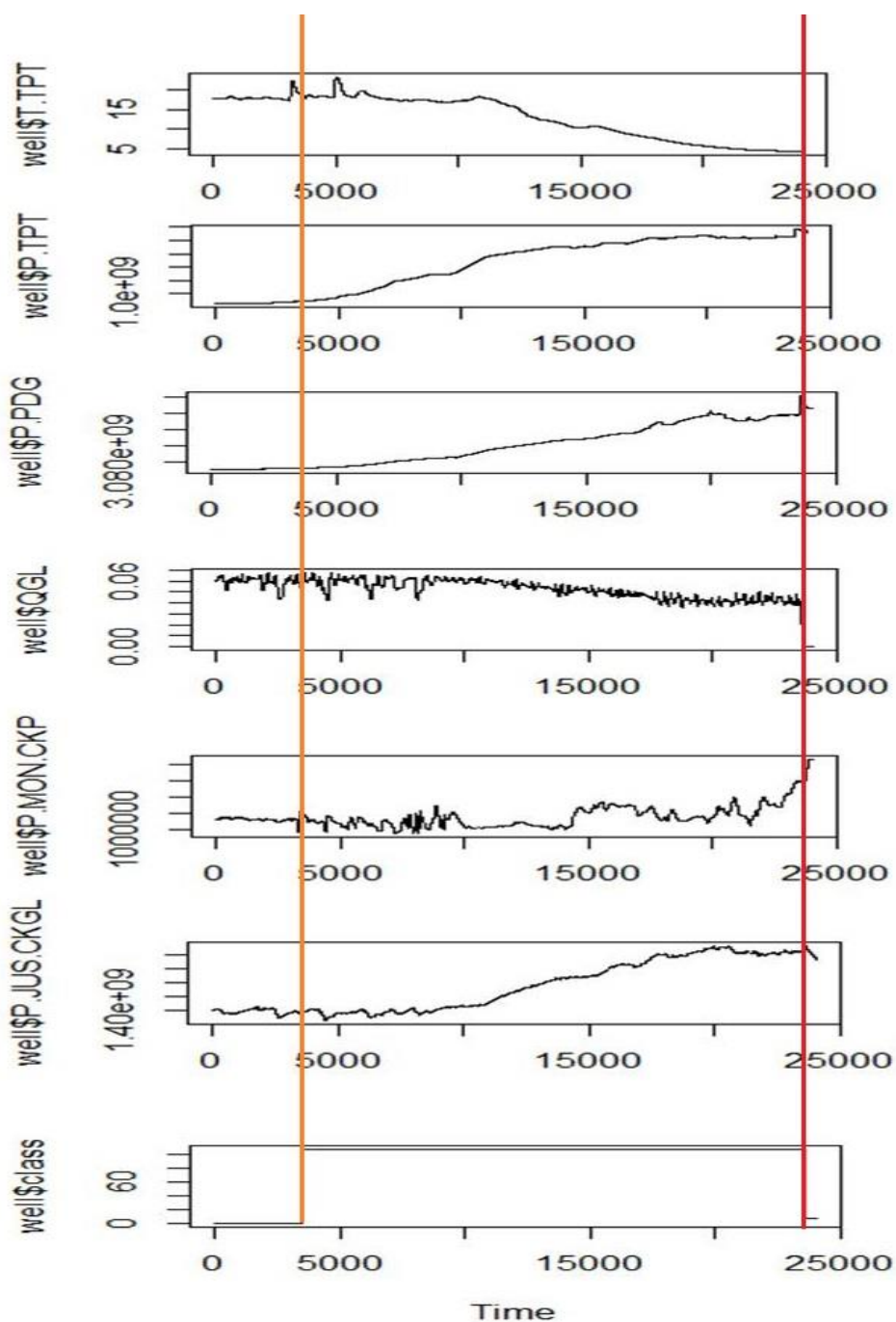


Figura 18- Comparação temporal das variáveis de produção do poço 21. A linha laranja indica o início do transiente e a linha vermelha indica a instalação da falha.

Nos poços contendo problemas de inscrustação na válvula de produção, o poço 01 e 06 apresentam valores nulos de vazão de *gas lift* e não possuem o sensor de temperatura à jusante da válvula de *gas lift*. Além disso, o poço 06 possui valores negativos de pressão no fundo do poço, enquanto o poço 01 apresenta valores nulos para essa mesma variável. Sendo assim, P.PDG, QGL e T.JUS.CKGL não serão contemplados

na análise dos poços. As informações estatísticas de ambos os poços são detalhadas nas Figuras 19 e 20.

timestamp	P.TPT	T.TPT
Length:43962	Min. :13286420	Min. :117.6
Class :character	1st Qu.:13342222	1st Qu.:117.8
Mode :character	Median :13406110	Median :118.0
	Mean :13444689	Mean :117.9
	3rd Qu.:13550625	3rd Qu.:118.1
	Max. :13685410	Max. :118.2
P.MON.CKP	T.JUS.CKP	P.JUS.CKGL
Min. :5257381	Min. :70.64	Min. :3383878
1st Qu.:5399837	1st Qu.:71.84	1st Qu.:3384159
Median :5481520	Median :73.12	Median :3384697
Mean :5503621	Mean :72.76	Mean :3384736
3rd Qu.:5609895	3rd Qu.:73.66	3rd Qu.:3385250
Max. :5752159	Max. :74.30	Max. :3385812
class		
Min. : 0.00		
1st Qu.:107.00		
Median :107.00		
Mean : 92.61		
3rd Qu.:107.00		
Max. :107.00		

**Figura 19- Estatística descritiva do poço 01 para problema de incrustação na válvula de produção**

timestamp	P.TPT	T.TPT
Length:68908	Min. :21125440	Min. :117.3
Class :character	1st Qu.:21193100	1st Qu.:117.4
Mode :character	Median :21206400	Median :117.5
	Mean :21202353	Mean :117.5
	3rd Qu.:21219140	3rd Qu.:117.6
	Max. :21246300	Max. :117.7
P.MON.CKP	T.JUS.CKP	P.JUS.CKGL
Min. :12032370	Min. :64.73	Min. :1627220
1st Qu.:12162090	1st Qu.:65.08	1st Qu.:2131432
Median :12205060	Median :65.44	Median :2402581
Mean :12196863	Mean :65.44	Mean :2363510
3rd Qu.:12235630	3rd Qu.:65.68	3rd Qu.:2625865
Max. :12354990	Max. :66.27	Max. :2842802
class		
Min. : 0.00		
1st Qu.:107.00		
Median :107.00		
Mean : 97.47		
3rd Qu.:107.00		
Max. :107.00		

**Figura 20- Estatística descritiva do poço 06**

As Figuras 19 e 20 apresentam o sumário estatístico dos poços 01 e 06, respectivamente. O tamanho do conjunto de dados para problemas de incrustação é muito maior que o de formação de hidratos. Observa-se que o patamar de pressão do poço 06 é maior que o do poço 01, além disso a variação de temperatura é menor em casos de

incrustação do que em formação de hidratos. A variação de temperatura tanto na árvore de natal quanto à jusante da válvula de produção é menor no poço 06 que no poço 01.

Na comparação temporal das variáveis do poço 01 na Figura 21, é possível observar que a pressão à montante da válvula de produção apresenta frequentes oscilações em seus valores. Conforme a produção deixa a condição normal, a variável apresenta uma tendência crescente, alcançando o pico com a entrada na classe de falha.

Todavia, a pressão à jusante válvula de *gas lift* apresenta uma queda, desde a condição normal da produção do poço, e em algum momento do transiente há a tendência de aumento da pressão, mas esta não volta aos valores obtidos na classe normal. Já a pressão na árvore de natal demora mais a caracterizar uma situação de transiente. A região normal e transiente próxima da linha laranja são similares. No entanto, com o passar do tempo, observa-se um aumento na pressão.

Em relação à temperatura na árvore de natal e à jusante da válvula de produção, percebe-se que há uma leve diminuição. Na árvore de natal a queda não chega a 1°C, enquanto na válvula de produção a variação é de cerca de 2°C, que não é muito se comparado às variações encontradas em poços com formação de hidratos.

O tamanho do conjunto de dados relativos ao poço 01 é maior que o dos poços vistos anteriormente. Com isso, a linha temporal mostrada nos gráficos acaba sendo comprimida, dificultando assim uma análise mais profunda usando aspectos visuais. Na região à direita da linha vermelha, onde já está instalado o problema de incrustação, não é possível verificar o progresso do comportamento das variáveis. O poço 06 só apresenta a condição normal e o transiente e, mesmo apresentando apenas duas classes, possui mais dados que o poço 01. Ou seja, o poço 06 possui mais horas de monitoramento que o poço 01 no banco de dados desenvolvido por Vargas *et al* (2019).



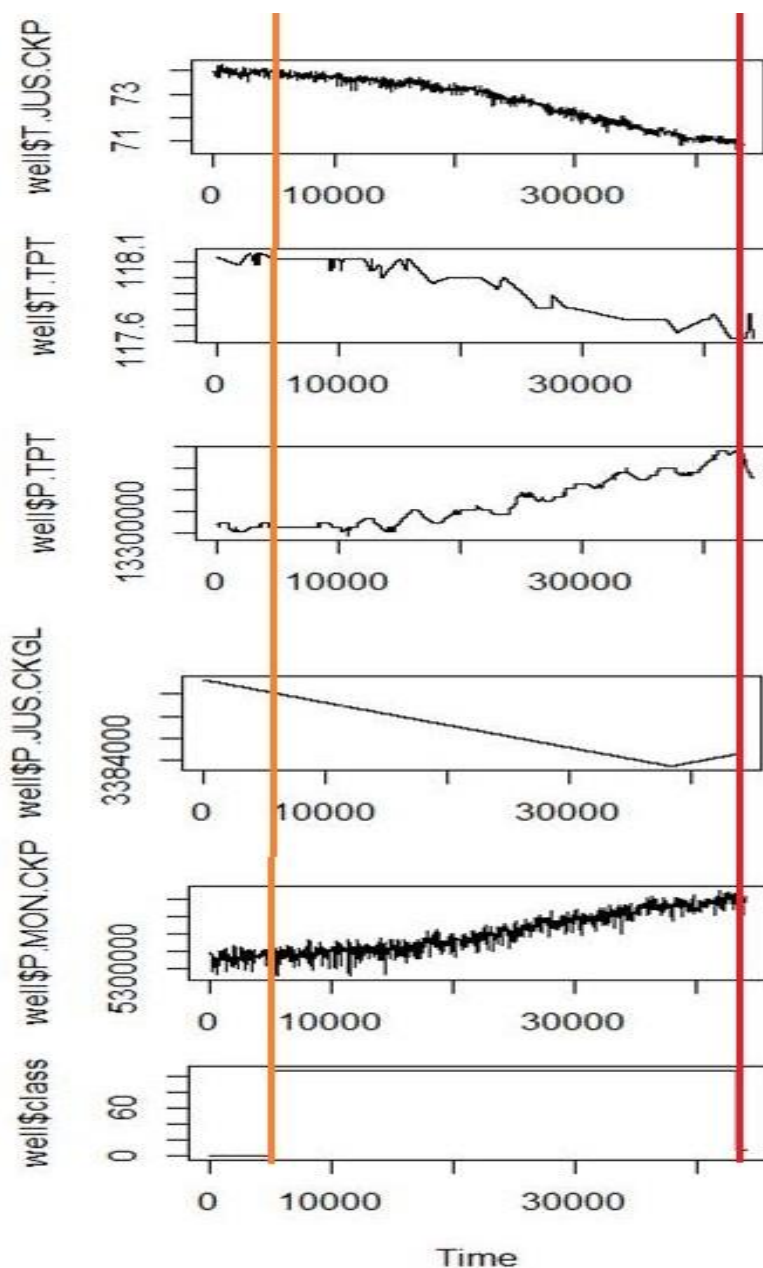


Figura 21-Comparação temporal entre as variáveis do poço 01. A linha laranja indica o início do transiente e a linha vermelha indica a instalação da falha.

Para fins de comparação do estado de transiente, também foi realizado a análise temporal do poço 06. A Figura 22 mostra que no poço 06 a frequência de oscilação da pressão à montante da válvula de produção é menor que no poço 01 e em ambos os poços a pressão nessa válvula aumenta gradativamente. A pressão na árvore de natal também apresenta aumento, porém, passada parte inicial do transiente, a pressão passa a oscilar em torno de um patamar e não aumentar gradativamente como no poço 01.

As variáveis temperatura na árvore de natal e à jusante da válvula de produção apresentam comportamentos similares nos dois poços, mostrando uma leve queda em

seus valores. A única variável que apresentou comportamento oposto ao observado no poço 01 foi a pressão à jusante da válvula de *gas lift*. Nesse poço, a partir do início da classe normal, a pressão apresenta aumento e esse comportamento perdura ao longo de todo o transiente. Enquanto no poço 01, essa variável apresentou queda.

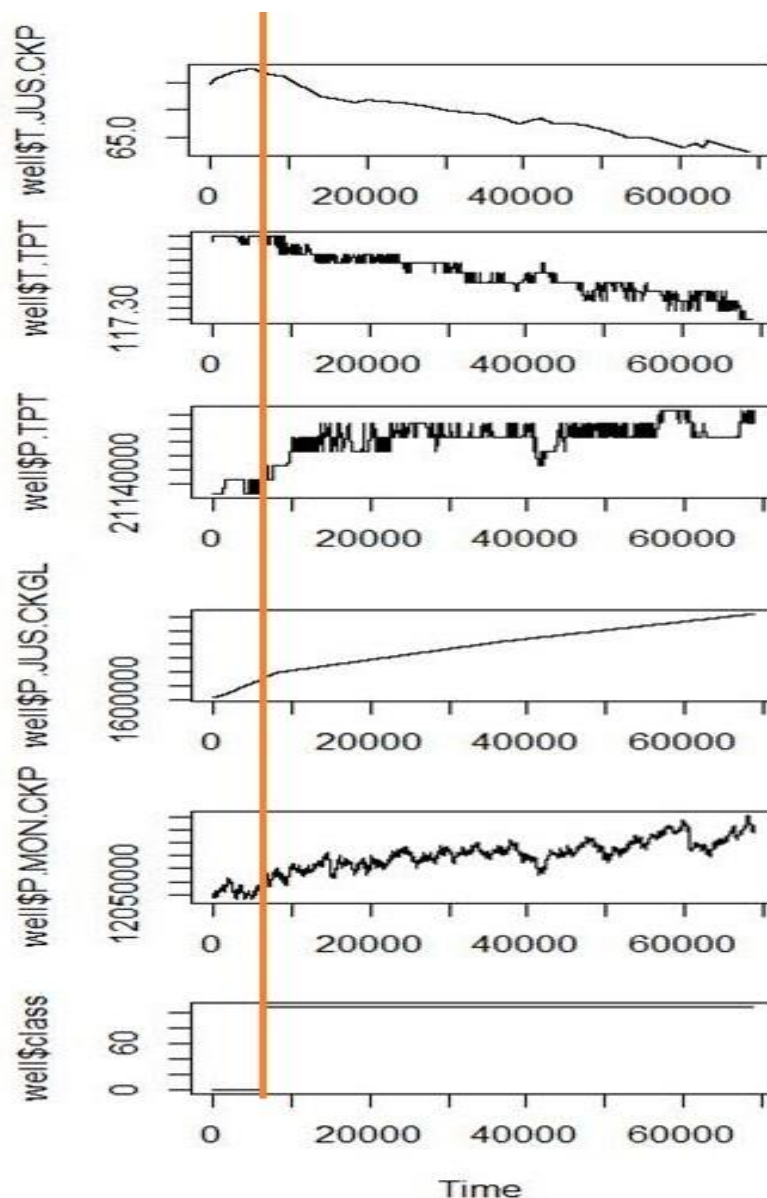


Figura 22- Comparação temporal entre as variáveis do poço 06, onde a linha laranja indica o início do transiente.

Por meio da análise temporal dos poços que apresentaram formação de hidratos e dos poços com problemas de incrustação foi possível observar como as variáveis se comportam no desenvolver dessas falhas. Para resumir os resultados obtidos nesta seção, foi montada a Tabela 5 mostra a amplitude de variação, correspondendo à diferença entre o maior e o menor valor de cada uma das variáveis estudadas.

**Tabela 5- Resumo da variação de amplitude de cada variável em cada poço analisado, usando a diferença entre o maior e o menor valor disponível.**

Variável	Formação de hidrato		Incrustação	
	Poço 20	Poço 21	Poço 01	Poço 06
T.JUS.CKP (°C)	Não possui	Não possui	-3.66	-1.54
T.JUS.CKGL	Não possui	Não possui	Não possui	Não possui
T.TPT (°C)	-57.17	-19.03	-0.6	-0.67
P.TPT (Pa)	$+1.15 \times 10^9$	$+1.13 \times 10^9$	$+3.99 \times 10^5$	$+1.2 \times 10^5$
P.PDG (Pa)	$+0.15 \times 10^9$	$+0.024 \times 10^9$	Não possui	Não possui
P.MON.CKP (Pa)	$+2.21 \times 10^6$	$+2.25 \times 10^6$	$+4.95 \times 10^5$	$+3.23 \times 10^5$
P.JUS.CKGL (Pa)	$+0.19 \times 10^9$	$+0.27 \times 10^9$	- 0.19	+0.27
QGL (m <sup>3</sup> /s)	-0.06	-0.07	Não possui	Não possui

Como mostra a tabela 5, poços distintos apresentam efeitos parecidos quando apresentam formação de hidrato. Em ambos os poços houve queda na temperatura na árvore de natal e aumento da pressão na linha de produção. Esse comportamento se repete nos poços com incrustação. No entanto, a redução da temperatura não é tão pronunciada como durante a ocorrência de hidrato. Em relação aos poços 01 e 06, houve apenas uma variável que não seguiu a mesma tendência em ambos os poços. A pressão à jusante da válvula de *gas lift* apresentou comportamento distinto para cada poço.

A análise temporal permitiu identificar o que ocorre com dados de pressão e temperatura no intervalo disponível para estudo. Como o conjunto de dados usado é extenso, alguns padrões e informações podem ter sido suprimidos na visualização da série temporal completa, como foi mostrado nas Figuras 17, 18, 21 e 22. Com os próximos experimentos será possível extrair ainda mais informações dos conjuntos de dados.

### 5.1.2. Experimento 1

O primeiro experimento do bloco 1 consiste em aplicar os algoritmos de classificação num conjunto dados do poço 20. O conjunto de treino e teste foram separados em 70% e 30%, respectivamente. Após a extração das instâncias nulas e NAs (not a number), foram aplicadas as técnicas de classificação detalhadas no capítulo 3.

- **GMM:** O modelo de misturas gaussianas aplicado no poço 20 obteve o F1-score de 99.23%. A classe “transiente” demandou uma quantidade significativamente maior de componentes de mixtura para englobar os padrões observados dentro dessa classe. A Figura 23 mostra a matriz de confusão gerada a partir da comparação entre os resultados preditos pelo modelo e a marcação das classes. Na parte superior da figura observa-se a marcação referência, ou seja, a marcação que está presente nos dados originais. Na lateral estão as classes preditas pelo modelo. Sendo assim, percebe-se que o modelo apresentou excelentes resultados. As classes normal e falha foram classificadas perfeitamente, enquanto apenas 16 instâncias que era “transientes” foram classificadas como falha já instalada.

Prediction	Reference		
	0	8	108
0	54	0	0
8	0	326	16
108	0	0	4284

**Figura 23-Matriz de confusão do modelo gerado pelo GMM**

- **Naive Bayes:** O modelo gerado para o classificador naive Bayes envolve gerar distribuições de densidade para cada classe disponível e variável analisada. Sendo assim, o algoritmo gerou um gráfico para cada um dos sensores estudados. Em cada gráfico há três distribuições de densidade correspondentes às classes normal, transiente e falha, como exemplifica a Figura 24 mostra um dos gráficos gerados pelo modelo. A distribuição escolhida no algoritmo foi a *kernel density estimation* por ser mais generalizável. A Tabela 6 apresenta os resultados obtidos da aplicação desse algoritmo para cada variável de forma resumida, apresentando a média de cada uma das classes. O *F1-score* obtido pelo classificador foi de 97.66% o que mostra a boa performance do algoritmo. Como é apresentado na matriz de confusão na Figura 25, as classes normal e de falha foram classificadas sem erros, mas o transiente foi marcado como falha em 44 instâncias.

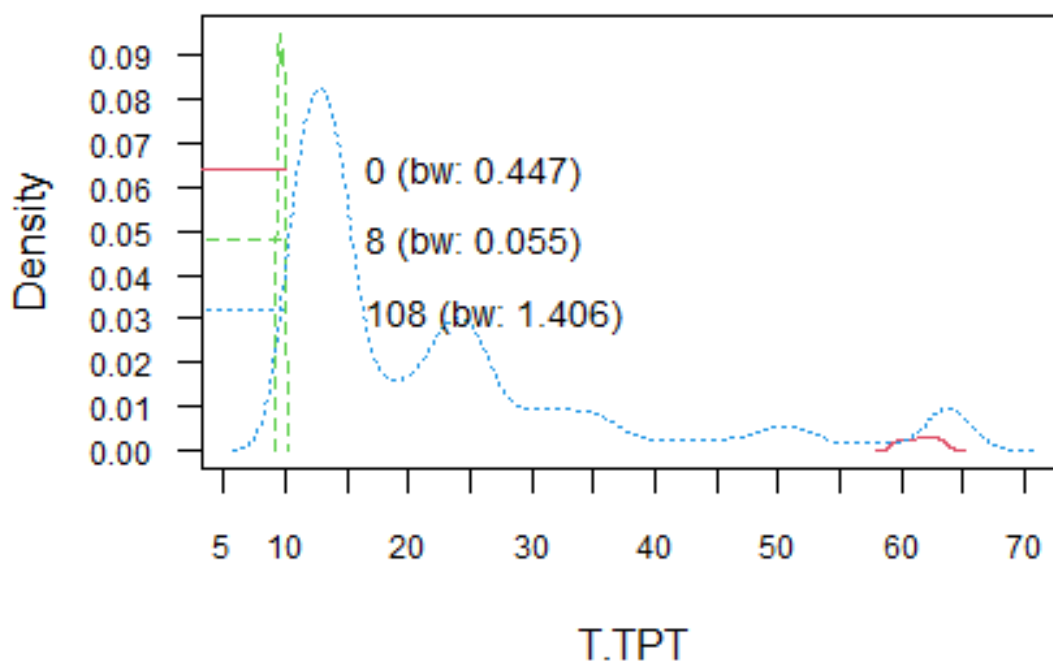


Figura 24- Exemplo de gráfico gerado pelo algoritmo naive Bayes

Tabela 6 – Resumo dos resultados obtidos com o classificador naive Bayes para cada sensor do poço 20.

	NORMAL	TRANSIENTE	HIDRATO
P.PDG	$2.159 \times 10^9$ Pa	$2.232 \times 10^9$ Pa	$2.306 \times 10^9$ Pa
P.TPT	$6.791 \times 10^8$ Pa	$12.51 \times 10^8$ Pa	$17.76 \times 10^8$ Pa
T.TPT	61.50°C	38.25°C	9.65°C
P.MON.CKP	$1.612 \times 10^6$ Pa	$2.046 \times 10^6$ Pa	$3.149 \times 10^6$ Pa
P.JUS.CKGL	$1.548 \times 10^9$ Pa	$1.614 \times 10^9$ Pa	$1.677 \times 10^9$ Pa
QGL	$0.0567 \text{ m}^3/\text{s}$	$0.029 \text{ m}^3/\text{s}$	$0.0002 \text{ m}^3/\text{s}$

	Reference		
Prediction	0	8	108
0	37	0	0
8	0	316	44
108	0	0	4283

Figura 25- Matriz de confusão do naive Bayes.

- Adaboost: o *script* usado para a aplicação do adaboost possui poucos parâmetros a serem estabelecidos. Foi escolhido 5 iterações para rodar o algoritmo de boosting pois como o conjunto de treino é extenso, ele demanda certo tempo de processamento. Com esse número de iterações já foi possível obter um resultado satisfatório para o modelo e não aumentou o tempo de processamento. Com essa técnica de classificação foi possível obter identificar todas as três classes perfeitamente, como mostra a matriz de confusão da Figura 26, e por consequência obteve-se um *F1-score* de 100%.

	Reference		
Prediction	0	8	108
0	37	0	0
8	0	316	0
108	0	0	4327

Figura 26- Matriz de confusão do modelo gerado pelo Adaboost.

- Árvore de decisão: A árvore de decisão, assim como o Adaboost, proporcionou a classificação sem erro das três classes presentes no conjunto de dados. A figura 28 representa a árvore de decisão gerada pelo conjunto de treino. A variável considerada mais importante, ou seja, com maior ganho de informação, para a classificação do conjunto de treino foi o P.PDG. A partir dele, exclusivamente, o algoritmo conseguiu distinguir as três classes corretamente. A matriz de confusão mostrada na Figura 27, corrobora a boa performance do algoritmo que apresentou o *F1-score* de 100%.

	Reference		
Prediction	0	8	108
0	37	0	0
8	0	316	0
108	0	0	4327

Figura 27-Matriz de confusão da árvore de decisão para problemas de formação de hidrato.

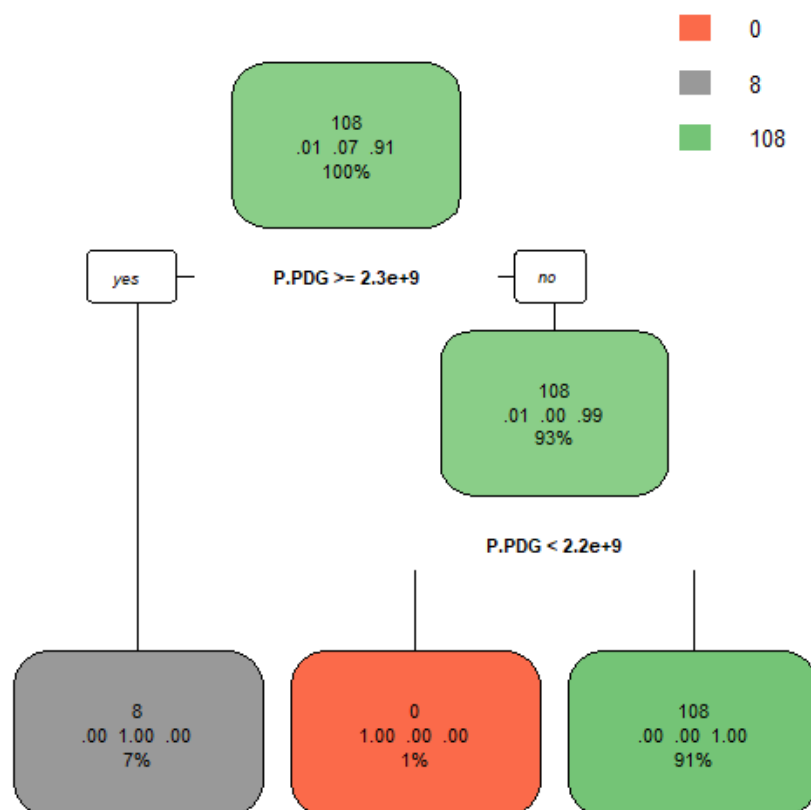


Figura 28-Árvore de decisão para formação de hidratos no poço 20

- *Random forest*: como uma única árvore de decisão já apresentou *F1-score* de 100%, o algoritmo de *random forest* também obteve o mesmo valor usando 300 árvores. Na árvore foi apenas apresentado os ramos necessários para a classificação, ou seja, somente com a informação de pressão no fundo do poço já foi possível obter resultados satisfatórios. No entanto, através do *decrease mean gini* calculado, denominado como “*importance*” no output do algoritmo, é possível determinar a importância de cada variável para a classificação da variável “classe”. A pressão no fundo do poço, como mostrado na árvore de decisão, é a que possui maior importância na determinação da classe. Temperatura na árvore de natal e vazão de *gas lift* também possuem influência para a classificação final.

	MeanDecreaseGini
P. PDG	522.23774
P. TPT	91.24687
T. TPT	420.05499
P. MON. CKP	31.19491
P. JUS. CKGL	241.53154
QGL	420.76681

**Figura 29- Importância de cada variável para classificação da variável alvo no conjunto de treino do poço 20**

### 5.1.3. Experimento 2

No segundo experimento realizado busca-se verificar se os algoritmos aplicados anteriormente, com o conjunto de treino baseado nas informações do poço 20, podem ser utilizados para detectar a formação de hidrato em um poço diferente. A análise dos poços mostrou que a temperatura e o patamar de pressão entre os poços são diferentes, porém os sensores disponíveis nesses poços são os mesmos. Após a normalização dos dados, o modelo baseado no poço 20 foi testado nos dados, também normalizados, do poço 21. A seguir é apresentada o desempenho de cada algoritmo de classificação, usando como conjunto de teste o poço 21 (Figura 30).

Reference Prediction    0    8    108 0    172    0    0 8    0    1109    49 108    1    0    14271	<b>GMM</b> F1-score = 99.12%
Reference Prediction    0    8    108 0    173    0    0 8    0    1109    140 108    0    0    14180	<b>Naive Bayes</b> F1-score = 97.86%
Reference Prediction    0    8    108 0    173    0    0 8    0    1109    0 108    0    0    14320	<b>Adaboost</b> F1-score = 100%
Reference Prediction    0    8    108 0    173    0    0 8    0    1109    0 108    0    0    14320	<b>Árvore de decisão</b> F1-score = 100%
Reference Prediction    0    8    108 0    173    0    0 8    0    1109    0 108    0    0    14320	<b>Random forest</b> F1-score = 100%

**Figura 30- Resultado da aplicação dos modelos treinados com dados do poço 20 para classificação do poço 21.**



Os modelos treinados a partir dos dados do poço apresentaram excelente desempenho ao classificar o poço 21, como mostra a Figura 30. O pior *F1-score* obtido foi de 97,86% pelo *naive* Bayes, onde 140 instâncias que pertenciam à classe “transiente” foi classificado como hidrato. Dessa forma o algoritmo está antecipando o transiente e o considerando como falha já instalada. Os demais algoritmos apresentaram resultados tão bons quanto os obtidos no conjunto de teste do poço 20. Esses resultados mostram que mesmo para um poço com características diferentes do utilizado para treinar os modelos foi possível detectar o estado transiente e a classe de falha por formação de hidrato. Sendo assim, considerando o banco de dados publicado por Vargas *et al* (2019), há a possibilidade de usar poços distintos para treinamento e teste de modelos, fazendo assim com os modelos gerados possam ser usados para detecção de formação de hidratos em mais de um poço.

#### 5.1.4. Experimento 3

O terceiro experimento desse bloco utiliza dados do poço 01 para treinar e testar os modelos gerados pelas ferramentas de aprendizado de máquina. A seguir serão apresentados os resultados obtidos a partir dos dados de incrustação na válvula de produção.

- GMM: este modelo apresentou um *F1-score* de 91,55%. Diferentemente do que ocorreu nos dados de formação de hidratos, o modelo gerado pelo GMM não obteve muito sucesso em distinguir o estado transiente do estado normal. Nesse modelo, o início da falha seria detectado de forma tardia em relação ao modelo obtido com dados de hidrato. A transição entre a classe “normal” e a classe “transiente” foi pior classificada que a transição da classe “transiente” para a classe “falha”. Ainda assim, os resultados obtidos pelo modelo são satisfatórios visto que apenas uma pequena parcela dos dados está fora da diagonal principal da matriz de confusão mostrada na Figura 31.

Prediction	Reference		
	0	7	107
0	1562	0	389
7	0	263	73
107	0	0	10901

Figura 31- Matriz de confusão relacionada ao GMM para o poço 01

- *Naive Bayes*: o algoritmo do naive Bayes obteve resultados melhores que o GMM, como mostra a Figura 32. O *F1-score* foi de 94,62%, o modelo errou mais instâncias da classe “transiente”, classificando-as como “normal”, porém distinguiu melhor a classe “transiente” da classe de “falha” porém acertou mais que As variações identificadas em cada variável como representado de forma resumida na Tabela 7 mostram-se mais suaves que as presentes em problemas de hidrato. A Tabela 7 apresenta a média das variáveis em cada classe presente no conjunto de dados. A variação de temperatura na árvore de natal, por exemplo, é insignificante.

Tabela 7- Resumo dos resultados obtidos com o classificador naive Bayes para cada sensor do poço 01

	NORMAL	TRANSIENTE	INCRUSTAÇÃO
P.TPT	13.32 x10 <sup>6</sup> Pa	13.48 x10 <sup>6</sup> Pa	13.59 x10 <sup>6</sup> Pa
T.TPT	118.1°C	117.9°C	117.7°C
T.JUS.CKP	73.92°C	72.43°C	70.86°C
P.MON.CKP	5.35x10 <sup>6</sup> Pa	5.50x10 <sup>6</sup> Pa	5.66x10 <sup>6</sup> Pa
P.JUS.CKGL	3.38 x10 <sup>6</sup> Pa	3.38x10 <sup>9</sup> Pa	3.38x10 <sup>9</sup> Pa

Prediction	Reference		
	0	7	107
0	1562	0	450
7	0	263	8
107	0	0	10905

Figura 32- Matriz de confusão referente ao algoritmo naive Bayes usado nos dados do poço 01

- *Adaboost*: este classificador aprendeu resultados excelentes também para dados de incrustação. O conjunto de treino do poço 01 é maior do que o referente ao poço 20, sendo assim o algoritmo apresentou maior tempo de processamento que

nos experimentos anteriores. O *F1-score* foi de 100%, a classificação pelo algoritmo de *boosting* continua a apresentar resultados promissores como mostra a Figura 33.

	Reference		
Prediction	0	7	107
0	1562	0	0
7	0	263	0
107	0	0	11363

Figura 33- Matriz de confusão referente ao algoritmo Adaboost usado nos dados do poço 01

- Árvore de decisão: o modelo obtido pela aplicação da árvore de decisão utilizou as variáveis P.JUS.CKGL e T.JUS.CKP para particionar o conjunto de treino. P.JUS.CKGL apresentou pouca variação em relação às médias das classes “0”, “107” e “7” como foi abordado na tabela 7. Por essa razão essa variável é utilizada mais de uma vez ao longo da árvore representada na Figura 34, o mesmo ocorre com T.JUS.CKP. O modelo apresentou *F1-score* de 97.80%, que é um bom resultado. De forma geral, como elucidado na Figura 35, o modelo errou aproximadamente a mesma quantidade de instâncias cuja classificação era “7” e o modelo considerou como “107”, estado de transiência, e instâncias com classificação “107” que foram consideradas falha já instalada

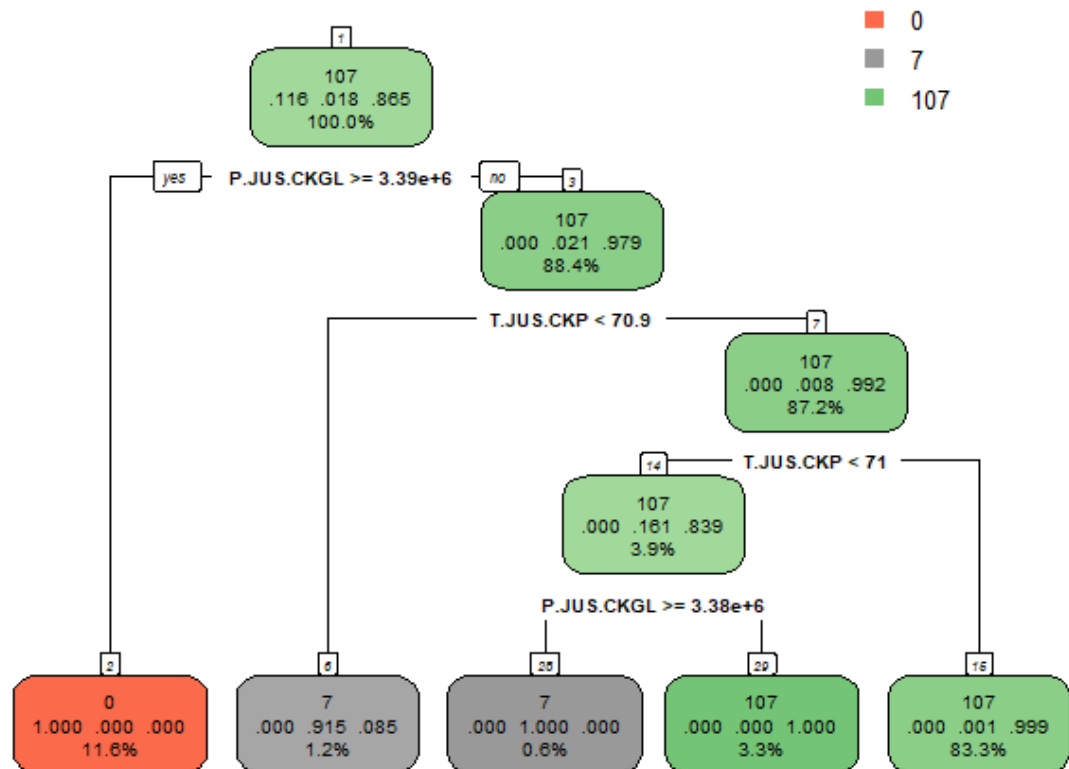


Figura 34- Árvore de decisão do poço 01

	Reference		
Prediction	0	7	107
0	1562	0	0
7	0	247	18
107	0	16	11345

Figura 35- Matriz de confusão do modelo gerado pela árvore de decisão

- *Random forest*: diferentemente do que ocorreu durante o experimento 2, a aplicação do algoritmo de *random forest*, no conjunto de dados relativo a problema de incrustação, pode mostrar como esse classificador atua de forma mais eficiente que uma única árvore. Em comparação com *F1-score* de 97,80% da árvore de decisão, a random forest obteve *F1-score* de 100% com menos de 25 árvores, sendo assim sua matriz de confusão é igual a da Figura 33. A importância das variáveis pode ser visualizada a seguir na Figura 36, que corrobora a disposição das variáveis na árvore de decisão. P.JUS.CKGL e T.JUS.CKP são variáveis mais importantes na determinação da classe das instâncias.

	MeanDecreaseGini
P.TPT	163.1360
T.TPT	896.5376
P.MON.CKP	260.5071
T.JUS.CKP	1953.2032
P.JUS.CKGL	4028.7478

**Figura 36- Importância de cada variável para classificação da variável alvo no conjunto de treino do poço 01**

#### 5.1.5. Experimento 4

Da mesma forma que foi realizado o experimento 2, porém utilizando os conjuntos de dados normalizados referentes à incrustação, esse experimento visa verificar a aplicabilidade dos modelos gerados por um poço para detectar a condição normal e transiente de outro poço. O poço 1 possui todas as três classes analisadas, enquanto o poço 06 usado para testar o modelo gerado pelo poço 1 possui apenas as classes normais e transientes. Como é da natureza da matriz de confusão ser uma matriz quadrada e o *F1-score* contabiliza os elementos fora da diagonal principal, foi necessário retirar a classe 7 para que tanto o conjunto de treino quanto o conjunto de teste estivessem com a mesma dimensão.

Os resultados obtidos neste experimento estão dispostos na Figura 37. O desempenho dos algoritmos GMM, *adaboost*, árvore de decisão e *random forest* foi consideravelmente inferior na classificação do poço 6. Somente o classificador naïve Bayes apresentou resultados positivos. Esse modelo foi capaz de identificar grande parte da condição transiente e normal do poço 6, e errou ao atrasar uma pequena parte do transiente, ou seja, o modelo classificou como normal os dados referentes ao estado transiente. De forma que ou os algoritmos não conseguiram classificar corretamente o estado normal do poço, considerando essa classe como transiente ou eles não conseguiram classificar o estado transiente, identificando este como estado normal. Sendo assim, diferente do que ocorreu no experimento 2 onde os modelos gerados por um poço foram capazes de classificar um poço distinto, para casos de incrustação na válvula de produção o mesmo não se confirmou. O conjunto de treino obtido pelos dados do poço 1 não obteve bons resultados na classificação do poço 6, exceto pelo naïve Bayes.

Reference			<b>GMM</b> F1-Score= 47.80%
Prediction	0	107	
0	8	0	
107	6127	62773	

Reference			<b>Naïve Bayes</b> F1-Score= 87.55%
Prediction	0	107	
0	6133	3470	
107	2	59303	

Reference			<b>Adaboost</b> F1-Score= 4.22%
Prediction	0	107	
0	3038	62773	
107	3097	0	

Reference			<b>Árvore de decisão</b> F1-Score= 4.22%
Prediction	0	107	
0	3038	62773	
107	3097	0	

Reference			<b>Random forest</b> F1-Score= 4.22%
Prediction	0	107	
0	3038	62773	
107	3097	0	

**Figura 37- Resultado da aplicação dos modelos treinados com dados do poço 1 para classificação do poço 6.**

#### 5.1.6. Discussão dos resultados dos experimentos do bloco 1

O bloco 1 apresentou, primeiramente, como os sensores de pressão, temperatura e vazão de *gas lift* variavam durante a ocorrência de hidratos e incrustação. A análise dos poços mostrou quais as variáveis estavam disponíveis nos conjuntos de dados de cada poço, apresentou o resumo estatístico de cada poço e uma análise temporal para verificar se os poços apresentavam comportamentos semelhantes para formação de hidratos e incrustação. Para os poços com dados de formação de hidratos, apenas as variáveis P.JUS.CKP e P.JUS.CKGL apresentaram comportamentos distintos durante a fase de transiência nos poços. Já para os poços com problema de incrustação, a única variável com comportamento divergente foi P.JUS.CKP.

A Tabela 5, obtida pelo cálculo da amplitude de variação de cada uma das variáveis envolvidas, compara as variações nos valores computados pelos sensores analisados para as duas anomalias estudadas neste bloco. Observando a temperatura na árvore de natal, é possível verificar que quando há a formação de hidratos a queda na temperatura é muito mais acentuada que durante a formação de incrustação; como já era de se esperar visto que baixas temperaturas são um fator-chave para a hidratação da linha

de produção. Como ambas anomalias geram redução do diâmetro disponível ao escoamento, a grande parte dos sensores de pressão apresentam aumento em seus valores.

As Tabelas 6 e 7 compilam os resultados do algoritmo *naive* Bayes. Nas tabelas constam os valores médios de cada classe em cada variável. Durante a ocorrência de hidratos, a pressão no fundo do poço aumenta e na árvore de natal é registrado um aumento ainda mais significativo, seguido de uma brusca redução da temperatura na árvore de natal. A pressão à montante da válvula de produção também apresenta aumento significativo. Na linha de *gas lift*, há aumento da pressão e redução da vazão de gás injetado. Por outro lado, na ocorrência de incrustação, não há aumento significativo da pressão na árvore de natal e a queda de temperatura é de menos de 1°C. Na plataforma, a linha de produção apresenta queda de temperatura maior que a da árvore, chegando a aproximadamente 3°C. A pressão à montante do *choke* também aumenta. Já na linha de *gas lift* a uma variação muito baixa da pressão.

Os experimentos 1 e 3 mostraram que os algoritmos de classificação apresentaram excelentes resultados utilizando os dados dos poços. Com isso, é possível concluir que utilizar tais classificadores para detectar ocorrências de anomalias nesses tipos de dados é uma opção viável. O experimento 2, testou a possibilidade de usar um modelo treinado com dados de um poço para prever ocorrência de hidratos em outro; a classificação realizada pelos modelos apresentou resultados positivos. Dessa forma, baseado no banco de dados utilizado nesse trabalho, verificou-se que modelos de classificação de hidratos podem ser aplicados em mais de um poço. Todavia, o mesmo resultado não se confirmou para casos de incrustação onde apenas o modelo do *naive* Bayes obteve resultado superior a 80%. No geral, os modelos gerados pelo poço 1 não foram capazes de classificar os dados do poço 6 com *F1-score* satisfatório.

## 5.2. Bloco de experimentos *slugging* e instabilidade de fluxo

Para anomalias como *slugging* e instabilidade de fluxo, há disponível apenas um poço para estudo dessas anomalias junto à sua condição normal, o poço 01. O conjunto de dados mostra, com horas de diferença, a ocorrência de *slug* seguido por uma instabilidade de fluxo. Em outro diretório do banco de dados publicado por Vargas et al (2019), há dados referentes à condição normal do poço 01 aproximadamente um mês antes e cerca de mês depois da ocorrência das anomalias. Dessa forma, para análise

temporal, também será mostrado como os sensores do poço se comportaram em condições normais antes e depois da mudança no regime de escoamento.

Em seguida, após a análise temporal do poço, os dados referentes à *slugging* e instabilidade de fluxo serão concatenados em um único conjunto de dados para que a eficiência dos algoritmos de aprendizado de máquina seja testada quanto à diferenciação desses dois tipos de problemas de garantia de escoamento.

### 5.2.1. Experimento 1: análise temporal

Como haverá apenas um poço a ser analisado durante o decorrer dos experimentos deste bloco, não se mostrou necessário apresentar o sumário estatístico de modo a mostrar o patamar de pressão e temperatura do poço, dentre outros parâmetros. O experimento 1 consiste na aplicação de séries temporais nos dados de *slugging* e instabilidade de fluxo do poço, como há apenas um poço com dados de anomalias e normalidade com intervalos de tempo próximos será feita uma análise temporal em sequência para cada variável.

Sumarizando as informações do poço 01 nos conjuntos de dados referentes à situação normal antes da ocorrência de *slugging*, durante *slugging* e instabilidade de fluxo, e retorno à situação normal foi possível construir a tabela 8 baseada na média de cada variável observada.

**Tabela 8 - Média das variáveis observadas em cada situação estudada para o poço 01**

	<b>Normal antes</b>	<b>Slugging</b>	<b>Instabilidade</b>	<b>Normal depois</b>
<b>P.PDG (MPa)</b>	0	0	0	0
<b>P.TPT (MPa)</b>	13.890229	14.062885	14.057236	14.182689
<b>T.TPT (°C)</b>	117.6	117	117	116.5
<b>P.MON.CKP(MPa)</b>	5.840973	6.029063	6.039816	6.084081
<b>T.JUS.CKP (°C)</b>	70	69.26	69.10	68.13
<b>T.JUS.CKGL (°C)</b>	Não possui	Não possui	Não possui	Não possui
<b>P.JUS.CKGL(MPa)</b>	3.402194	3.245779	3.246440	3.416956
<b>QGL (m<sup>3</sup>/s)</b>	0	0	0	0

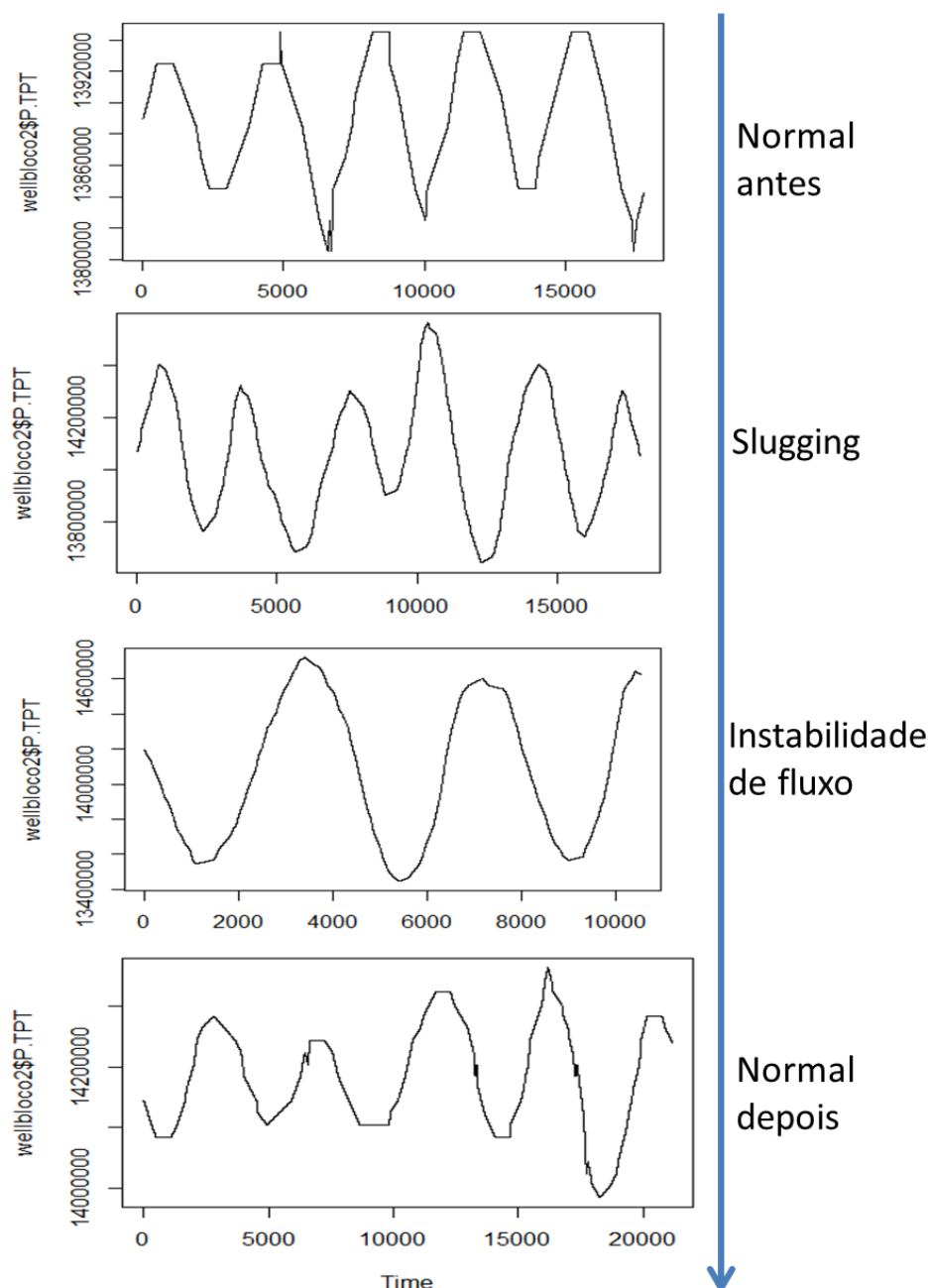
O poço 01 não possui dados de temperatura à jusante da válvula de *gas lift* e apresenta valores nulas para pressão no fundo do poço e vazão de *gas lift*. A tabela 8



mostra que após a ocorrência de *slugging* e instabilidade de fluxo o padrão de normalidade do poço é alterado, os sensores de pressão apresentam patamar mais alto enquanto a temperatura apresenta pequena diminuição. É possível que o período de um mês não seja o suficiente para que o poço retome aos valores observados antes das anomalias, porém, estudando dados de dois meses após a ocorrência dessas anomalias, também verificou-se que o padrão normal não retomou aos valores obtidos anteriormente.

Pelo fato de haver apenas um poço em estudo será apresentado o comportamento individual de cada sensor nas quatro situações. Dessa forma é possível ver com mais detalhes os impactos gerados pelo *slugging* e pela instabilidade de fluxo durante a produção.

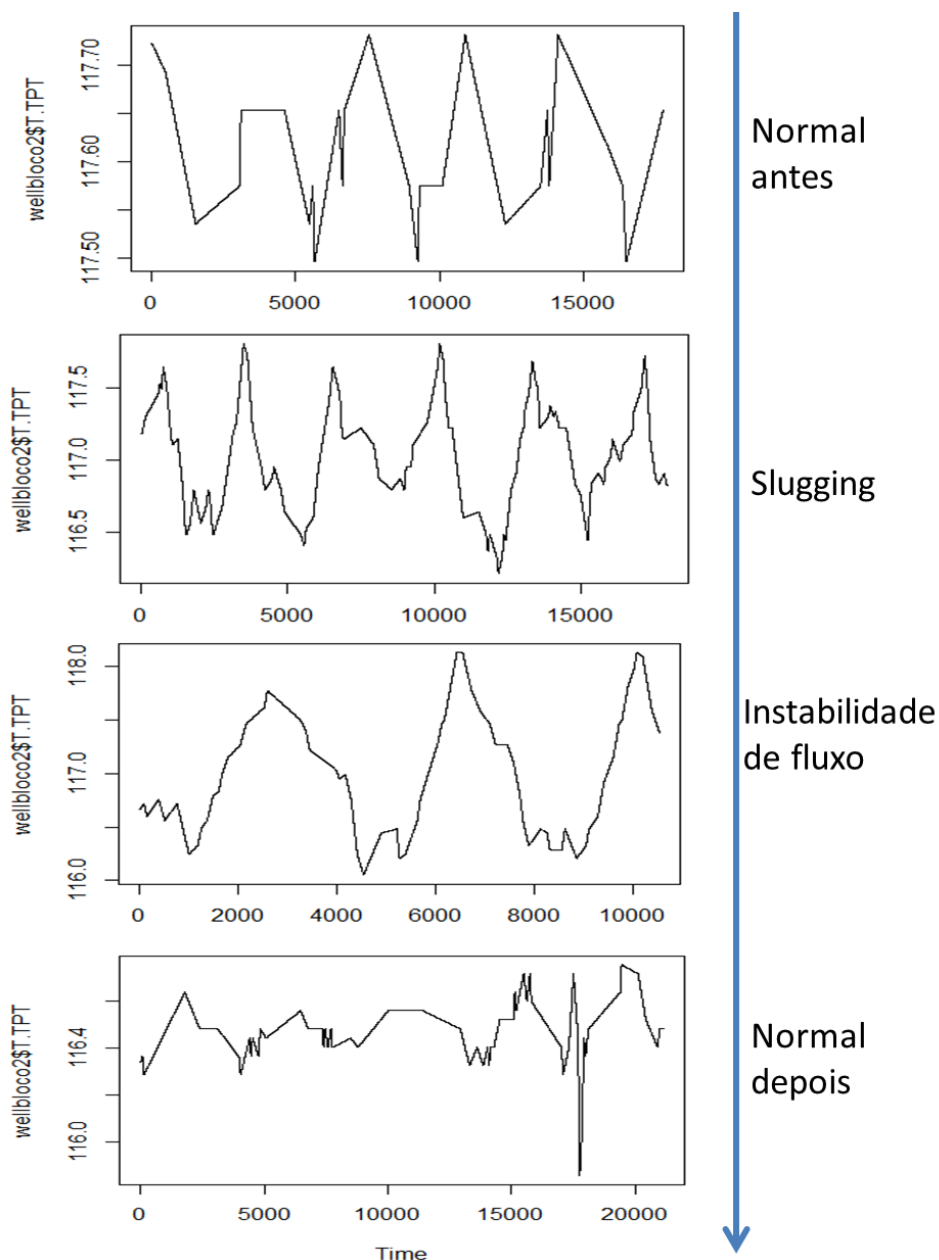
Primeiramente, começando pelo sensor de pressão na árvore de natal, como está disposto na figura 38, observa-se que no mês anterior ao aparecimento das anomalias o sensor indicava variações regulares de pressão. A pressão oscilava entre 13,8 MPa e 13,9 MPa, nas cristas e vales observados na Figura 38 percebe-se que há um platô mostrando uma estabilização nesses valores por um curto intervalo de tempo. Com o início de *slugging*, o patamar de pressão na árvore de natal muda, passando variar entre valores de 13,6 MPa e 14,5 MPa. A queda de pressão aumenta em decorrência das golfadas de gás alternadas a trechos de líquidos, o aumento da amplitude dos valores observados no sensor de pressão demonstra um comportamento periódico, já esperado em casos de *slugging*. A ocorrência de *slugging* é seguido por uma instabilidade no fluxo, caracterizada ainda pelo aumento da amplitude entre os valores máximos e mínimos de pressão, onde são medidos valores entre 13,4 MPa e 14,7 MPa. O comportamento periódico é muito similar ao observado durante a ocorrência de *slugging*. Na retomada à normalidade após cerca de um mês depois das anomalias acontecerem, o patamar de pressão não retoma exatamente ao mesmo patamar da situação “normal antes”, porém a amplitude da queda de pressão diminui em relação as situações de falha. A pressão oscila entre 13,9 MPa e 14,3 MPa e, por breves períodos, há estabilização da pressão assim como na situação normal anterior aos eventos.



**Figura 38- Sensor de pressão na árvore de natal (P.TPT) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo**

Ainda na árvore de natal, o comportamento da temperatura medida nesse ponto da arquitetura *subsea* é apresentado na Figura 39. A variável T.TPT apresenta baixa variação, na situação “normal antes” a diferença entre a temperatura máxima e mínima é de apenas 0,2°C, e a média é de 117,6°C. Durante o *slugging* essa diferença aumenta para 1,6°C, ou seja, há um aumento na amplitude entre os picos e vales mostrados no segundo gráfico da Figura 39, já a temperatura média é de 117°C. Com a instabilidade de fluxo, a diferença aumenta para 2°C e a temperatura média permanece em 117°. Por fim, após um

mês, a diferença entre temperaturas máximas e mínimas diminui passando a ser 1,1°C e temperatura média cai para 116,5°C.



**Figura 39- Sensor de temperatura na árvore de natal (T.TPT) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo**

Já nos equipamentos de superfície da unidade de produção, como mostra a Figura 40, a pressão à montante da válvula de produção atinge valores de 5,7 a 5,9 MPa durante a condição normal de produção. Durante a ocorrência de *slugging*, há um aumento nos picos de pressão à montante dessa válvula, a pressão passa a oscilar de 5,7 a 6,2 MPa e

apresentar períodos maiores. Passa a fase de *slugging*, durante a instabilidade de fluxo a amplitude entre os valores mínimos e máximos é um pouco acima da observada durante o *slug*, podendo chegar a 6,3 MPa. O caráter periódico não é proeminente nessa fase, diferentemente do que ocorre durante o *slugging*. Por último, na retomada à normalidade do poço, ainda que o patamar de pressão ainda seja superior, o comportamento da variável retoma ao padrão normal obtido antes das anomalias.

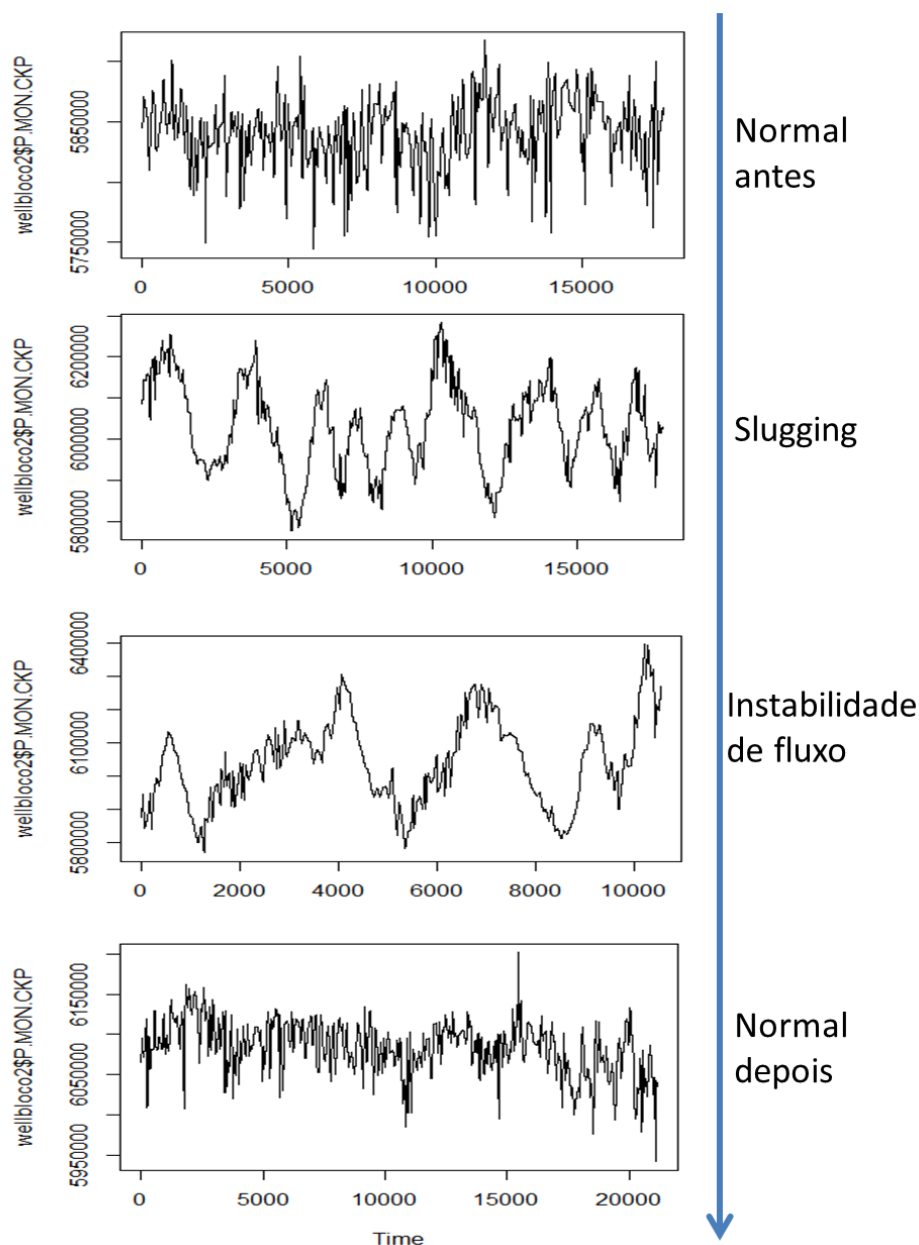
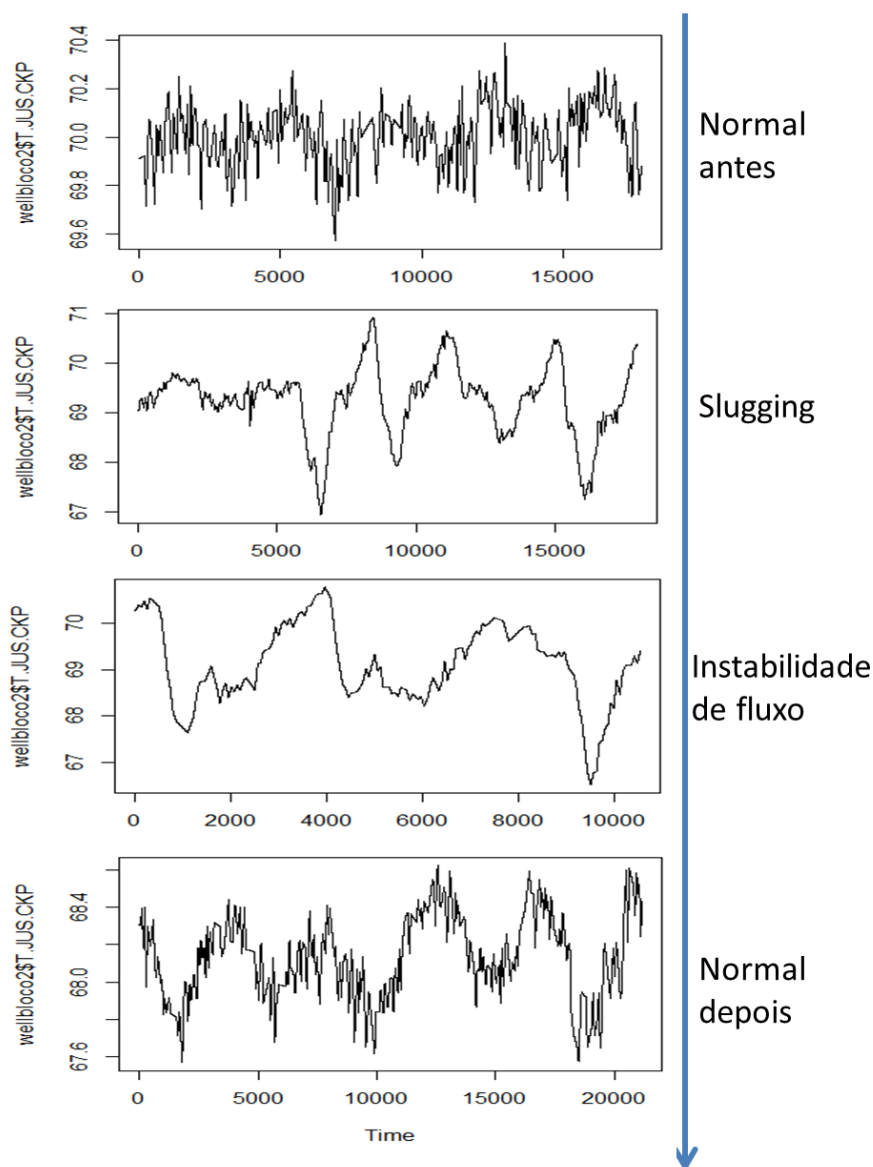


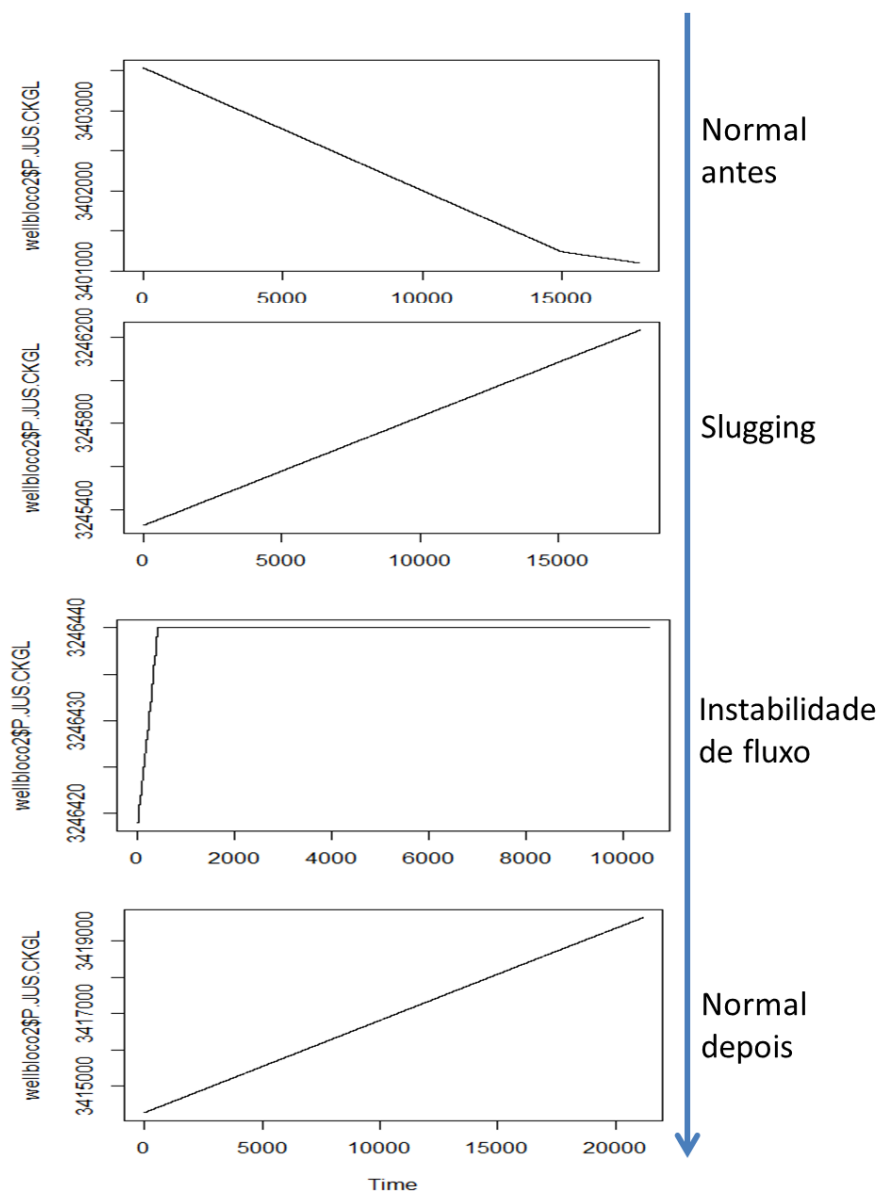
Figura 40- Sensor de pressão à montante da válvula de produção (P.MON.CKP) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo

Ainda na linha de produção, o sensor de temperatura à jusante da válvula de produção tem seu comportamento detalhado na Figura 41. Durante a produção normal do poço a variação de temperatura é de menos de 1°C. Com o surgimento do *slugging*, a temperatura passa a variar mais, com valores de 66,93°C a 70,92°C. Durante a instabilidade de fluxo, a temperatura permanece variando entre os mesmos valores, sem diferenças significativas em relação ao comportamento apresentado durante o evento de *slugging*. Por fim, a redução da faixa de variação da temperatura, voltando a variar cerca de 1°C (indo de 67,57 °C a 68,62°), sinaliza a volta da condição normal do poço.



**Figura 41- Sensor de temperatura à jusante da válvula de produção (T.JUS.CKP) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo**

Ainda analisando as variáveis da linha de *gas lift*, é possível observar através da Figura 42 que a pressão à jusante da válvula de injeção de *gas lift* não apresenta comportamento oscilatório como as variáveis estudadas anteriormente. Desde a situação normal a variação é pequena, cerca de 0,002434 MPa. Durante o *slugging*, a variação pressão é ainda menor, porém o patamar de pressão médio passa de 3,4 MPa para 3,2 MPa. Esse patamar se mantém durante a instabilidade de fluxo, e, após a normalização da produção, retoma aos valores próximos observados antes da ocorrência das falhas. Diferentemente do ocorrido nos sensores de pressão na linha de produção, a Figura 42 indica que variável P.JUS.CKGL não é impactada pelas oscilações geradas pelo *slugging* e instabilidade de fluxo como os sensores de pressão da linha de produção.



**Figura 42- Sensor de pressão à jusante da válvula de injeção de gas lift (P.JUS.CKGL) antes, durante e depois da ocorrência de slugging e instabilidade de fluxo**

O comportamento das variáveis disponíveis do banco de dados apresentado anteriormente foram tratadas de forma isolada, buscando se entender como cada uma delas se comportava individualmente durante os cenários propostos. Com a análise isolada perde-se sensibilidade de comparação com as demais variáveis, sendo assim se torna difícil comparar a variação da pressão, por exemplo, na árvore de natal e no *topside* da plataforma. Tendo em vista que o profissional responsável pelo monitoramento do poço pode ter acesso a mais de uma variável ao mesmo tempo faz se necessário também a análise comparativa entre a variação de pressão e de temperatura nos diferentes sensores

estudados. Dessa forma, torna-se visível como cada sensor é impactado pela ocorrência de *slugging* e instabilidade de fluxo.

Durante a produção normal do poço, é pequena a variação da pressão na árvore de natal. Já na válvula de *gas lift* essa variação é imperceptível em comparação aos demais sensores. Por outro lado, na válvula de produção, observa-se grande variabilidade na pressão, que não permanece fixa em um mesmo valor. A Figura 43 mostra que durante a produção normal do poço os sensores de pressão da árvore de natal e na linha de *gas lift* apresentam poucas flutuações em seus valores quando comparado à variação observada na válvula de produção.

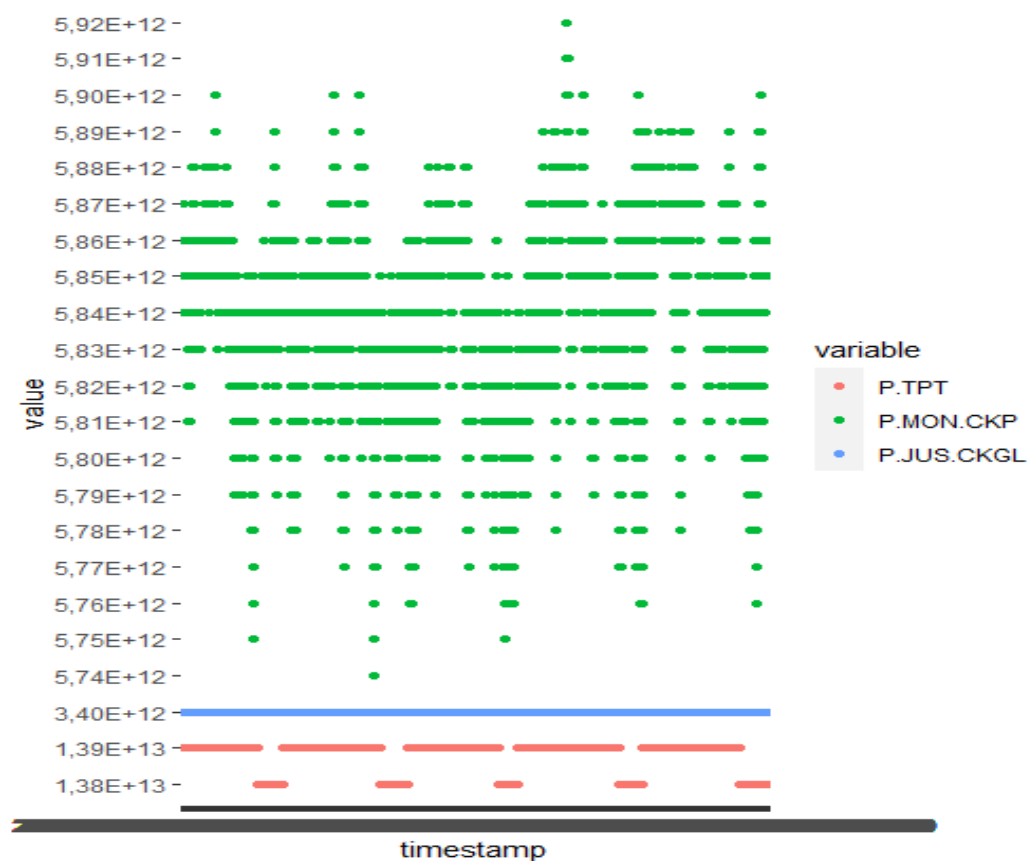


Figura 43- Sensores de pressão durante a produção normal do poço

Com o padrão de intermitência causado pela ocorrência de *slug*, a pressão na válvula de produção passa a apresentar picos de oscilação com maiores amplitudes. A pressão na árvore de natal também desenvolve comportamento periódico, característico desse regime de escoamento. Por outro lado, a pressão na linha de *gas lift* não aparenta ser afetada pela mudança no regime de escoamento, como indica a Figura 44.



Passadas algumas horas após o *slugging*, começa a fase de instabilidade de fluxo. Assim como na Figura 44, essa fase também é caracterizada por oscilações na pressão da válvula de produção e na árvore de nata. No entanto, a amplitude de variação da pressão nesses aumentam e o caráter periódico das oscilações ainda permanece, como pode ser observado na Figura 45.

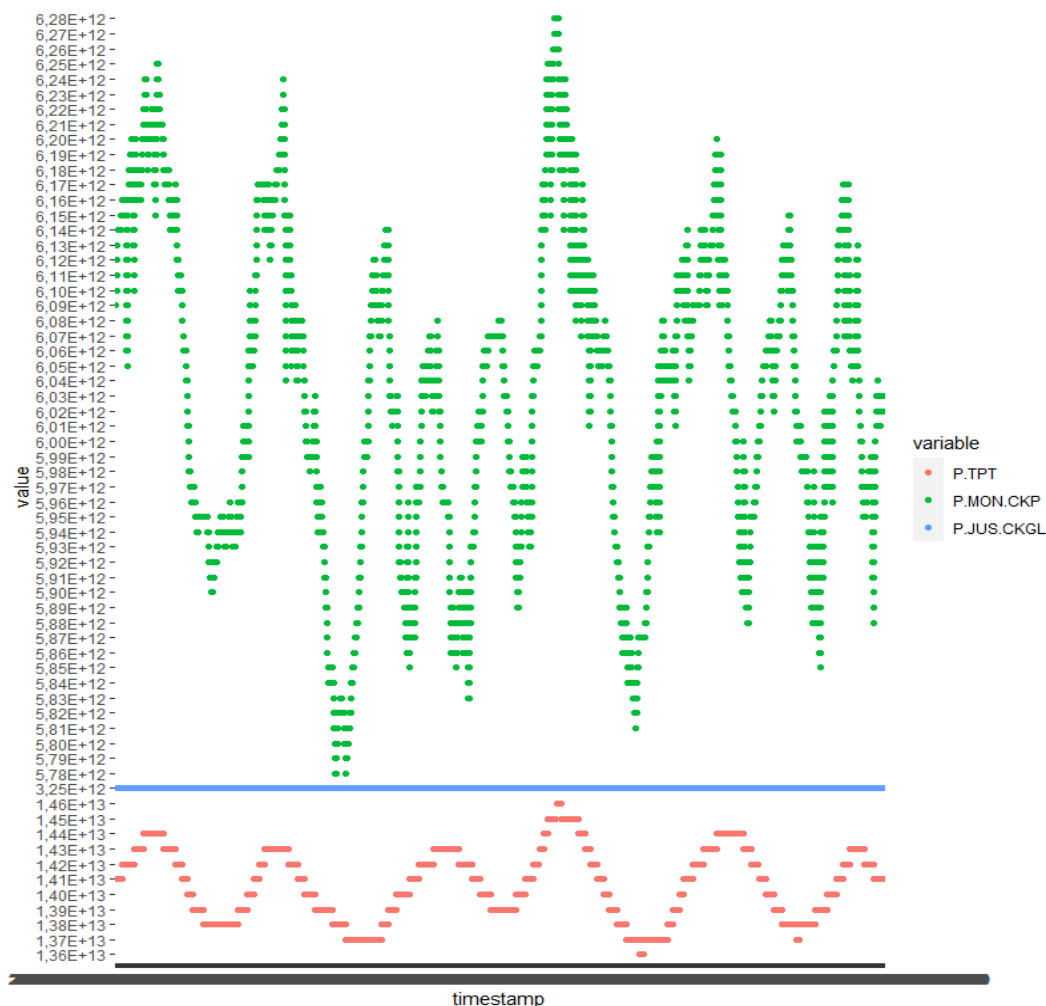


Figura 44- - Comportamento dos sensores de pressão durante slugging

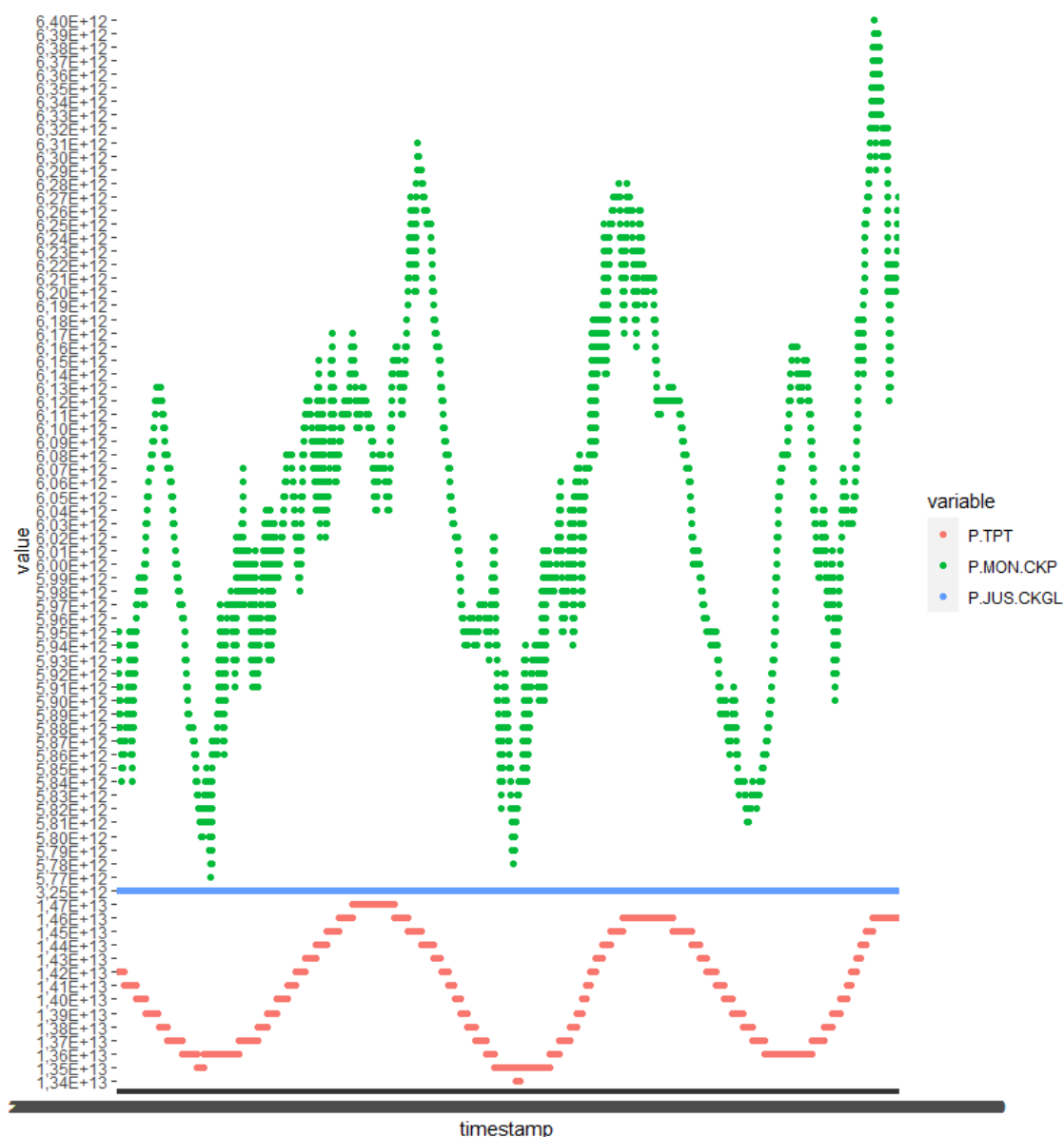
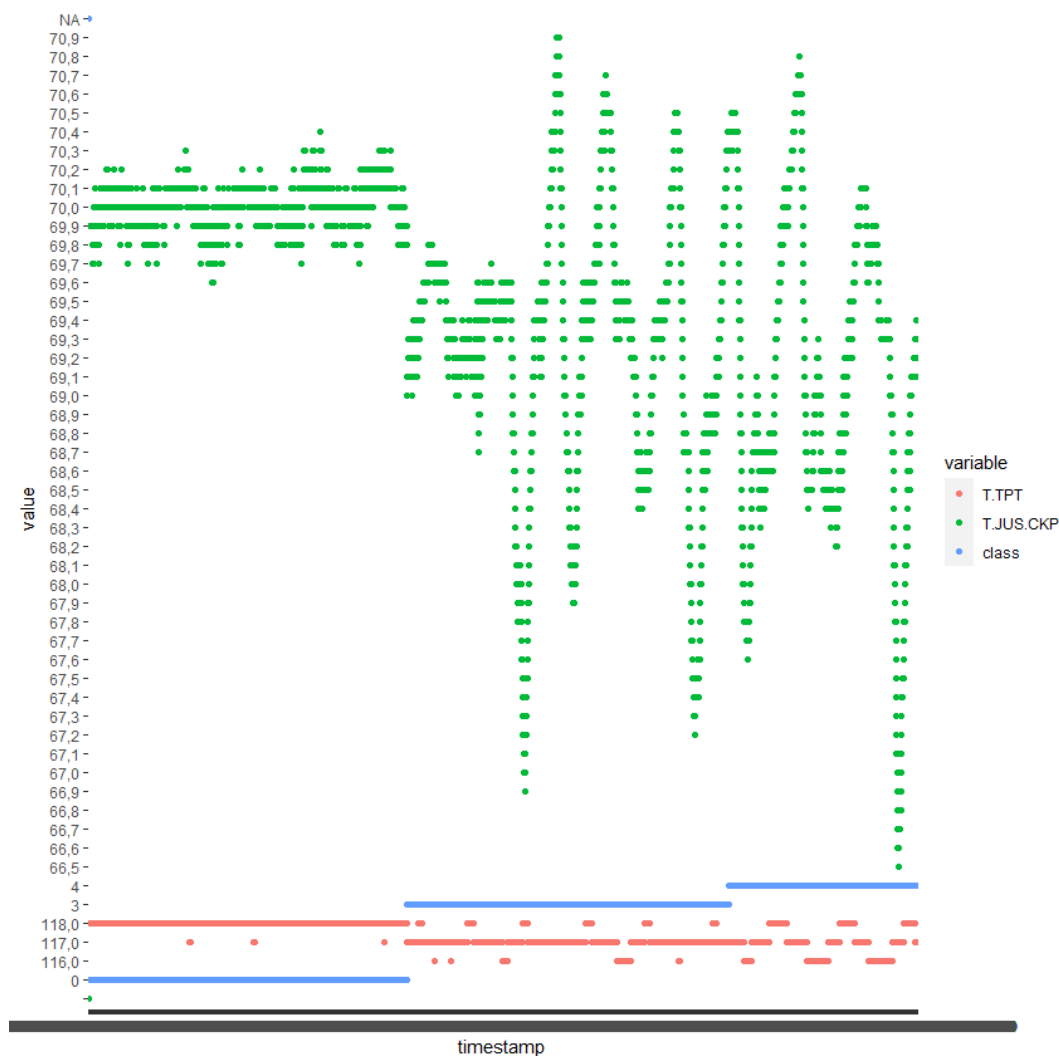


Figura 45- Comportamento dos sensores de pressão durante a instabilidade de fluxo

Contudo, diferente da pressão, a temperatura possui amplitude de variação numericamente menor. O que permite compilar o comportamento da temperatura na árvore de natal e na válvula de produção em um único gráfico sem que haja perda de escala. Nesse sentido, a Figura 46 mostra a partir da mudança na classificação dos dados, como os sensores de temperatura variam. Primeiro, há poucas variações na temperatura da árvore de natal e pequenas oscilação na temperatura à jusante da válvula de produção. Então com a mudança na classe, que sinaliza o início do *slugging*, há inicialmente queda na temperatura de ambos os sensores; seguido de flutuação nos valores principalmente na

válvula de produção. Com a mudança de classificação de *slugging* para instabilidade de fluxo, é possível observar que a temperatura na árvore de natal passa oscilar mais a medida que a temperatura à jusante da válvula de produção não apresenta mudança clara em seu comportamento.



**Figura 46-** Comportamento dos sensores de temperatura durante os cenários “normal”, “slugging” e “instabilidade de fluxo”.

Na análise temporal do poço 01 foi possível visualizar o desenvolvimento das variáveis de pressão e temperatura durante a produção normal do poço e durante a ocorrência de *slugging* e instabilidade de fluxo. As Figuras 38 a 42 mostram o comportamento temporal dos sensores disponível nesse poço, informação disponível durante o processo de monitoramento do poço. Todavia, as Figuras 43 a 46 apresentam de forma agrupada os sensores de pressão e temperatura. Sendo assim é possível comparar

como cada sensor reage durante as situações em análise, verificando qual a amplitude de variação entre os sensores de pressão e temperatura.

### 5.2.2. Experimento 2

O experimento 2 consiste no estudo da aplicação dos algoritmos de classificação num conjunto de dados contendo informações de *slugging* e instabilidade de fluxo no mesmo poço com a diferença de poucas horas entre a ocorrência dessas anomalias. Dessa forma, o novo conjunto de dados possui duas classes; 3 e 4 (*slugging* e instabilidade de fluxo, respectivamente).

Assim como nos experimentos do bloco 1, também foi separado 70% dos dados para treinar do modelo e 30% para testá-lo. A sequência de aplicação dos algoritmos também foi a mesma utilizada no bloco de experimentos anterior. De modo que a Figura 47 sumariza o resultado de cada um dos algoritmos, mostrando a matriz de confusão gerada pelo modelo e o *F1-score* obtido.

Reference Prediction    3    4 3 5394    75 4    0 3083	<b>GMM</b> F1-score: 99.06%
Reference Prediction    3    4 3 5394    0 4    0 3158	<b>Naive Bayes</b> F1-score: 100%
Reference Prediction    3    4 3 5394    0 4    0 3158	<b>Adaboost</b> F1-score: 100%
Reference Prediction    3    4 3 5394    0 4    0 3158	<b>Árvore de decisão</b> F1-score: 100%
Reference Prediction    3    4 3 5394    0 4    0 3158	<b>Random forest</b> F1-score: 100%

**Figura 47- Resultado da aplicação dos algoritmos de classificação nos dados de *slugging* e instabilidade de fluxo concatenados**

Na classificação de anomalias como *slugging* e instabilidade de fluxo, os algoritmos de aprendizado de máquina apresentaram excelentes resultados. O pior desempenho foi observado pelo algoritmo GMM com *F1-score* de 99,06%, onde algumas instâncias referentes à instabilidade de fluxo foram classificadas como *slugging*. Todos os demais algoritmos foram capazes de diferenciar as anomalias em estudo.

Em particular, a árvore de decisão também possibilita a visualização dos nós usados para determinar a classificação dos dados. A Figura 48 mostra a árvore gerada pelo algoritmo, nela a variável P.JUS.CKGL foi suficiente para dividir o conjunto de dados. Na análise temporal essa variável não apresentou variação significativa quando comparada com os demais sensores de pressão (Figura 45), porém observando apenas o comportamento desse sensor é visível, na Figura 42, que há aumento da pressão quando o poço deixa de apresentar *slugging* e passa a sofrer com instabilidade de fluxo. Com isso, apenas com essa variável a árvore de decisão foi capaz de classificar corretamente todas as instâncias do conjunto de dados concatenados.

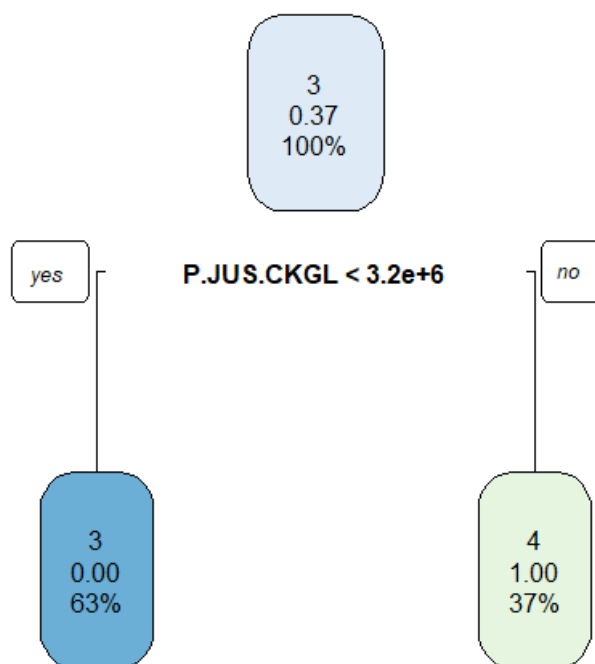


Figura 48- Árvore de decisão para classificação de dados de *slugging* e instabilidade

Como a árvore de decisão já obteve o melhor resultado possível, o algoritmo de *random forest* apenas se igualou em desempenho. Todavia, outro parâmetro fornecido por esse algoritmo que corrobora o exposto na Figura 48 é a importância de cada variável

na construção do modelo classificador. Na Figura 49 apresentada a seguir é possível observar que a importância da variável P.JUS.CKGL é muito superior que as demais, seguida pelas variáveis da árvore de natal.

	MeanDecreaseGini
P.TPT	818.05430
T.TPT	400.95457
P.MON.CKP	61.05923
T.JUS.CKP	183.44134
P.JUS.CKGL	7838.49222

**Figura 49- Importância das variáveis para a construção do modelo de classificação da random forest**

### 5.2.3. Experimento 3

O terceiro experimento desse bloco consiste em utilizar os modelos treinados a partir dos dados do poço 01 para detecção de *slugging* e instabilidade de fluxo em um conjunto de teste com dados de um poço diferente. Nesse caso, o poço 14, disponível no banco de dados públicos, apresentou ocorrência de instabilidade de fluxo seguida de *slugging* com intervalo de tempo menor que um dia entre a finalização da primeira e início da segunda. Com isso, os dados de ambas as anomalias foram concatenados em um único arquivo para ser posteriormente usado como conjunto de teste dos modelos de classificação gerados a partir do poço 01.

Vale ressaltar que a variável “*timestamp*”, referente à data e hora da ocorrência das anomalias, é eliminada durante o treinamento e testagem dos modelos de classificação. Por isso, o fato de a instabilidade de fluxo preceder a ocorrência de *slugging* ou o inverso, não deve ocasionar incoerência na análise.

O poço 14 também possui mais variáveis que o poço 01. Dessa forma, as variáveis que não são comuns aos dois poços não serão usadas, uma vez que os algoritmos de classificação estão limitados às variáveis disponíveis no momento da criação e treinamento do modelo.

Após realizada a exclusão das variáveis excedentes e a limpeza das variáveis restantes, os modelos de classificação foram aplicados no poço 14 e os resultados estão dispostos na Figura 50. Os resultados mostram que os modelos classificadores construídos com os dados do poço 01 não apresentaram bom desempenho na detecção de *slugging* e instabilidade de poços do poço 14. O melhor resultado obtido foi o do algoritmo GMM com *F1-score* de 41,66%, onde apenas a classe de *slugging* foi

classificada corretamente e como esta apresentava maior número de dados, o *F1-score* desse algoritmo foi superior aos demais.

Prediction	Reference			
	3	4		
3	17832	7142		
4	0	0		
			<b>GMM</b>	F1-score: 41.66%

Prediction	Reference			
	3	4		
3	0	0		
4	17832	7142		
			<b>Naive Bayes</b>	F1-score: 22.23%

Prediction	Reference			
	3	4		
3	0	0		
4	17832	7142		
			<b>Adaboost</b>	F1-score: 22.23%

Prediction	Reference			
	3	4		
3	0	0		
4	17832	7142		
			<b>Árvore de decisão</b>	F1-score: 22.23%

Prediction	Reference			
	3	4		
3	0	0		
4	17832	7142		
			<b>Random forest</b>	F1-score: 22.23%

**Figura 50-** Resultado obtidos pelos modelos criados a partir do poço 01 para classificação dos dados de *slugging* e instabilidade de fluxo no poço 14

Com exceção do algoritmo GMM, todos os outros modelos testados acertaram apenas a classe de instabilidade de fluxo e identificaram *slugging* como sendo instabilidade de fluxo. Essa anomalia pode ocorrer antes e desenvolver para *slugging* ou ocorrer depois do *slugging*. De maneira geral, o escoamento caracterizado por *slugging* gera instabilidade no fluxo do fluido produzido. Os resultados obtidos nesse experimento mostram que não é indicado usar os mesmos modelos para poços distintos. Por um lado, quando usados apenas dados do poço 01 para treino e teste dos modelos de classificação atingiu-se 100% de *F1-score*; por outro lado, esses mesmos modelos não mostraram bons resultados para classificação do poço 14.

#### 5.2.4. Discussão dos resultados dos experimentos do bloco 2

No primeiro bloco de experimentos as anomalias analisadas possuíam em comum a restrição do diâmetro disponível para escoamento da produção causada pela aglomeração de partículas sólidas. Neste segundo bloco de experimentos, ambas as anomalias apresentavam fatores que tornavam a produção de fluidos instável. No primeiro experimento desse bloco foi aplicada a técnica de séries temporais nos dados do

poço 01 que continha a condição normal do poço antes e depois da ocorrência das anomalias, bem como dados das anomalias em si. Como apenas um poço estava sendo analisado, optou-se por estudar como cada variável se comportava ao longo dessas quatro situações.

A Tabela 8 mostra os valores médios das variáveis em cada uma das situações propostas. Segundo os resultados dessa tabela, os sensores de temperatura mostram baixa amplitude de variação se comparado com anomalias como formação de hidrato e inscrustação. Os sensores de pressão apresentam maior variabilidade que a temperatura e, visivelmente, é possível observar mudanças nos valores de pressão em cada uma das situações.

Em seguida, a análise temporal de cada uma das variáveis foi realizada de modo a mostrar a amplitude de variação delas e o comportamento temporal no desenvolver de cada situação. As Figuras 38 a 42 mostram a variação nos sensores de pressão e temperatura no decorrer das situações de normalidade de falha que seriam observadas durante o monitoramento do poço. Contudo, o sensor de pressão da árvore de natal apresentou pequenos indícios que o poço estava saindo da condição normal. A Figura 38 mostra que entre a situação “normal antes” e “*slugging*” é possível observar apenas que deixa de haver um patamar máximo de pressão, passando a haver picos de pressão. Com a mudança para instabilidade de fluxo há mudança no período de oscilação da pressão nesse sensor e a volta à normalidade pode ser observada de forma clara. Nos demais sensores de pressão a alteração na condição de escoamento do poço se dá de forma mais clara, visualmente. Já o sensor de temperatura na árvore de natal não mostra um comportamento claro durante a mudança no estado do poço, enquanto o sensor de temperatura à jusante da válvula de produção, assim como os sensores de pressão, apresenta clara mudança de comportamento no decorrer das quatro situações.

Após realizada a análise temporal individual para cada variável, foram agrupados os sensores de pressão e temperatura separadamente. Os sensores de pressão, conforme mostram as Figuras 43, 44 e 45, foram analisados durante a condição normal do poço e durante a ocorrências das anomalias.

Na condição normal do poço, a pressão na à montante da válvula de produção apresenta comportamento difuso e a pressão na árvore de natal apresenta pequenas oscilações. Com o *slugging*, percebe-se que o comportamento difuso da variável P.MON.CKP passa a apresentar caráter mais oscilatório, assim como a variável P.TPT. O comportamento oscilatório da variável P.TPT apresenta maior período entre um pico e



outro em decorrência da instabilidade de fluxo. A pressão à jusante da válvula de injeção de *gas lift* apresenta variação desprezível quando comparada com os outros sensores, apesar de mostrar comportamento bem delineado durante cada uma das situações como mostra a Figura 42.

Os sensores de pressão foram concatenados em apenas um gráfico contendo as três situações de interesse (Figura 46). A transição da condição normal para *slugging* pode ser observada em ambos os sensores de temperatura, de forma mais clara no sensor de temperatura à jusante da válvula de produção. Durante a transição de *slugging* para instabilidade de fluxo, a variável T.TPT é a que detecta melhor que houve mudança de uma anomalia para outra.

No segundo experimento do bloco 2, foram aplicados os algoritmos de classificação de modo a detectar no poço 01 a ocorrência de *slugging* ou instabilidade de fluxo. Novamente, os modelos gerados pelos algoritmos apresentaram resultados excelentes na distinção dessas anomalias. O pior desempenho foi obtido pelo GMM que apresentou *F1-score* de 99.06% onde algumas instâncias que eram referentes à instabilidade de fluxo foram classificadas como sendo *slugging*. A árvore de decisão mostrou que a variável P.JUS.CKGL sozinha foi capaz de classificar corretamente os dados do poço 01, como foi corroborado pelo atributo “importância das variáveis” extraído como subproduto da aplicação do algoritmo *random forest*.

Usando os modelos gerados no experimento 2, no experimento 3 foram utilizados dados referentes ao poço 14 para testar os modelos na classificação de poços distintos do utilizado no treinamento dos modelos. Foram utilizadas as mesmas variáveis no teste do modelo, mesmo havendo maior número de variáveis disponíveis no poço 14. Os resultados dispostos na Figura 50 mostram que os modelos não foram capazes de distinguir as duas anomalias no poço 14. Sendo assim, os modelos de classificação para detecção de *slugging* e instabilidade de fluxo devem ser treinados e testados com dados de um mesmo poço.

# Capítulo 6

## Considerações finais

A pesquisa desenvolvida nesta dissertação aborda o uso de ferramenta de aprendizado de máquina para diferenciar problemas de garantia de escoamento como formação de hidratos, incrustação, *slugging* e instabilidade de poço a partir de dados reais disponíveis na literatura. Com base no comportamento das variáveis computadas por sensores localizados no fundo do poço, no leito marinho e na unidade de produção foi possível analisar como cada uma das anomalias estudadas impactavam as variáveis monitoradas do poço.

As questões propostas nessa pesquisa foram respondidas ao longo do capítulo 5, onde foram apresentados quais eram as diferenças entre os poços com diferentes anomalias, quais valores cada variável assume nos três estados estudados, quais sensores conseguem dividir melhor o conjunto de dados nesses estados e quais algoritmos obtiveram melhor desempenho em identificar essas condições para cada anomalia.

Para distinguir os problemas de garantia de escoamento que ocorreram nos poços disponíveis no banco de dados utilizado, foram formados dois grupos. Um relacionado a problemas que geram redução do diâmetro disponível ao escoamento e outro relacionado a instabilidades geradas durante a produção de fluidos. Assim, usando técnicas de aprendizado de máquina, foram criados modelos classificadores que dado um conjunto de dados reais foram capazes de distinguir condições normais de produção de condições onde houve ocorrência de anomalias. Os modelos de classificação, no caso do bloco 2, também foram capazes de identificar dados referentes a *slugging* e a instabilidade de fluxo com base no modelo treinado.

Por meio da compreensão da variação de pressão e temperatura ao longo da linha de produção, profissionais encarregados do monitoramento do poço conseguem detectar a ocorrência de anomalias. Algumas vezes, com a experiência de campo, também é possível identificar o tipo de falha que está se desenvolvendo no poço. Dessa forma, a pesquisa proposta nessa dissertação também apresentou o estudo das variáveis durante a ocorrências desses problemas de escoamento, mostrando seu comportamento temporal e os valores alcançados, de modo a gerar conhecimento para que outros profissionais possam adquirir uma fração da sensibilidade de análise que é obtida em campo, após anos observando o monitoramento de poços.

Os resultados dos experimentos do bloco 1 mostram que é possível diferenciar a condição normal, transiente e de anormal de um poço que desenvolveu hidrato ou incrustação. Como cada poço de petróleo e gás possui características únicas, os valores médios observados em cada classe podem variar de acordo com o poço. Portanto, ao comparar diferentes poços nos experimentos 2 e 4, foi necessário normalizar os dados de treino e teste a serem usados.

Ao utilizar as técnicas de aprendizado em um mesmo poço, foram obtidos resultados positivos na caracterização das classes normal, transiente e anormal. Na classificação dos poços com formação de hidrato, os algoritmos *adaboost*, árvore de decisão e *random forest* apresentaram o melhor desempenho obtendo um *F1-score* 100%. Já nos poços com incrustação, os melhores algoritmos foram, novamente, *adaboost* e *random forest*.

Ainda no bloco 1, verificou-se a possibilidade de utilizar modelos de classificação gerados a partir de um poço para classificar outros poços. Os modelos criados com os dados do poço 20 foram usados na classificação do poço 21, que também apresentava formação de hidrato, os resultados mostraram que mesmo com características diferentes, os algoritmos detectaram com bom desempenho as três classes presentes no poço 21. O mesmo não se observou para os poços com incrustação, onde o melhor resultado obtido foi de 87% de *F1-score* para apenas um dos algoritmos enquanto os demais apresentaram resultados inferiores a 50%.

Em relação ao estudo das variáveis, os resultados apresentados no bloco 1 mostraram que os dados das variáveis de pressão no fundo do poço, temperatura na árvore de natal e vazão de *gas lift* foram mais importantes para detectar se os dados eram correspondentes às classes normal, transiente ou falha por formação de hidrato. Para identificar a normalidade, o estado de transiência e falha por incrustação, por outro lado, foram as variáveis de pressão à jusante da válvula de *gas lift* e temperatura à jusante da válvula de produção que se destacaram como portadoras de informações para identificação das três classes que o poço poderia apresentar.

Nas matrizes de confusão elaboradas a partir da predição dos modelos e da classe de referência dos dados, ficou claro que a classe transiente é a mais difícil de ser classificada. Todos os modelos que não obtiveram *F1-score* de 100% apresentaram erros na classificação do transiente; ou os modelos antecipavam o transiente, classificando como normal dados da classe transiente, ou os modelos identificavam tardiamente que a

anomalia já estava desenvolvida, classificando como transiente os dados de falha instalada.

Os experimentos do bloco 2 foram realizados de maneira diferente. Como *slugging* e instabilidade de fluxo não apresentavam estado normal e transiente em seus arquivos, foi necessário compor um banco de dados contendo, além dos problemas de escoamento, períodos onde o poço estava com a produção normal. Com isso, o comportamento das variáveis foi estudado durante a fase normal antes da ocorrência das anomalias, *slugging*, instabilidade de fluxo e quando o poço retornou à normalidade.

A análise temporal dos sensores de pressão e temperatura mostrou que a pressão na árvore de natal e a pressão à montante da válvula de produção apresentam comportamento periódico tanto durante o *slugging* quanto durante a instabilidade de fluxo, dificultando a diferenciação desses problemas. Ao observar o comportamento individual dos sensores de pressão é possível ver que há sim mudanças que caracterizam a transição de *slugging* para instabilidade. A pressão à jusante da válvula de *gas lift* apresentou a maior distinção entre as situações analisadas, conforme mostra a Figura 42. A árvore de decisão e o algoritmo *random forest* destacaram essa variável como importante na classificação de *slugging* e instabilidade de fluxo, a pressão na árvore de natal foi a segunda variável mais impotante.

No bloco 2, os modelos de classificação foram aplicados para identificar *slugging* e instabilidade de fluxo. Os algoritmos usados apresentaram bons resultados na classificação desses problemas, mostrando que utilizar tais ferramentas na identificação de dados dessas anomalias, num mesmo poço, traz resultados positivos. Contudo, utilizar poços diferentes do usado para construir o modelo de classificação não gera resultados satisfatórios. Para poços diferentes os modelos identificaram apenas um tipo de anomalia, ou seja, as particularidades dos poços influenciam no comportamento dos sensores.

A presente dissertação apresentou, a partir de um conjunto de dados reais disponível na literatura, uma análise temporal dos sensores de produção, mostrando como é possível detectar anomalias como formação de hidratos, incrustação, *slugging* e instabilidade de fluxo, e a aplicabilidade de ferramentas de aprendizado de máquina para classificar dados de problemas de garantia de escoamento durante a produção de petróleo. Para melhor gestão de poços *offshore* é necessário entender os tipos de anomalias que ocorrem no poço e quais suas consequências no sistema de produção. Classificar os dados referentes a problemas de garantia de escoamento pode melhorar a compreensão a cerca dos impactos gerados nos sensores e auxiliar profissionais responsáveis pelo

monitoramento do poço no processo decisório, mostrando quando o poço como e quando o poço saiu da condição normal ou indicando se o problema se trata de *slugging* ou instabilidade de fluxo, por exemplo.

Os dados utilizados para construção dos modelos são estáticos, são recortes de períodos que foram previamente classificados pelos autores Vargas *et al.* Dados reais de poços não apresentam as classificações “normal”, “transiente” e “falha”, é preciso primeiro obter um conjunto de treino classificado para que os algoritmos de aprendizado de máquina classifiquem novos dados. Dessa forma, é imprescindível o fornecimento de dados de qualidade e uma boa classificação dos dados a serem usados no treinamento dos modelos, que irão conter as informações das classes alvo dos algoritmos de aprendizado de máquina.

### 6.1. Conclusões

Resumindo os pontos abordados na seção anterior, as conclusões dessa pesquisa podem ser dispostas em forma de tópicos. Com isso, os principais pontos a serem realçados são:

- O sensor de pressão localizado no fundo do poço o atributo que melhor classificou o conjunto de hidratos em “normal”, “transiente” e “falha”;
- O sensor de pressão localizado à jusante da válvula de *gas lift* e de temperatura à jusante da válvula de produção foram, respectivamente, os atributos que melhor classificaram os dados de incrustação na válvula de produção em “normal”, “transiente” e “falha”;
- A pressão à jusante da válvula de *gas lift* também foi o atributo que melhor distinguiu dados referentes a *slugging* severo e instabilidade de fluxo;
- Os algoritmos *adaboost*, árvore de decisão e *random forest* apresentaram os melhores resultados para dados de hidrato, incrustação, *slugging* e instabilidade de fluxo para dados de treino e teste oriundo de um mesmo poço;
- Os modelos de classificação construídos a partir dos dados normalizados de formação de hidrato do poço 20 foi capaz de classificar um poço diferente, também com dados normalizados;
- Os modelos de classificação construídos a partir de dados normalizados de incrustação no poço 01 não obteve bons resultados quando aplicados no poço 06. Apenas o algoritmo do *naive Bayes* apresentou *F1-score* acima de 80%.

O presente trabalho contém aspectos relacionados a área de garantia de escoamento e aprendizado de máquina. A literatura correspondente à garantia de escoamento mostra que hidratos de gás representam o maior desafio a ser combatido na produção de campos marítimos (SLOAN, 2009). Incrustação também consta como um problema significativo durante a produção de óleo e gás, tanto em terra quanto *offshore*. Nesse sentido, para hidratos há estudos para predição da formação de hidratos a partir de correlações envolvendo pressão, temperatura e outros parâmetros (CHAUDHARI; ZERPA; SUM, 2018; MAJID et al, 2018), assim como modelos de predição envolvendo aprendizado de máquina (GLÉNAT *et al*, 2015; REBAI *et al*, 2019). Assim como apontado pela literatura, os dados de pressão e temperatura na árvore de natal foram os mais determinantes para caracterizar se o poço estava produzindo normalmente ou com indícios de formação de hidratos.

No entanto, para incrustação, que também é um problema de interesse, é possível encontrar estudos mais diversificados, sobre incrustação em reservatórios, bombas, tubulações e em válvulas. Contudo, não há estudos com direcionamento abordado nesta dissertação que possam ser citados como forma de comparação dos resultados.

Por outro lado, para instabilidade de fluxo e *slugging* comprovaram-se os comportamentos periódicos e a variação de pressão nos sensores localizados na linha de produção causada pela ascensão da golfada de gás pelo *riser*. Assim como descrito nos trabalhos de Theyab (2018) e Taitel (1986).

## **6.2. Desafios futuros**

A grande questão a ser colocada é como seria a aplicação dos algoritmos usados nesta dissertação em dados medidos em tempo real. Um dos desafios da utilização de informações em tempo real é que os modelos precisam detectar o desenvolvimento de padrões e ocorrência de anomalias em um intervalo de tempo que possibilite a tomada de ações por parte dos profissionais de monitoramento. Ou seja, a detecção das anomalias deve ser realizada de forma preditiva. Os modelos apresentados nessa dissertação necessitariam de um conjunto de treinamento atualizado continuamente, conforme ele fosse perdendo a sensibilidade do que é o padrão de normalidade do poço. Além disso, os algoritmos que antecipam a condição transiente seriam mais indicados para detectar precocemente a mudança de *status* do poço.

Outras estruturas são necessárias na construção de um sistema de detecção em tempo real, como uma arquitetura para executar a análise dos dados dos sensores e todos os modelos de classificação, mostrando qual o do estado do poço naquele momento. A sensibilidade do modelo também deverá ser analisada por tal arquitetura, visto que os patamares de pressão e temperatura podem mudar com o tempo e, assim, os modelos também precisam ser adaptados.

O estudo da aplicação dessas ferramentas em dados reais sofre pela falta de oportunidade de se gerar dados reais de poço em tempo reais enquanto os modelos estão sendo desenvolvidos na fase de pesquisa. Sendo assim, o teste dos algoritmos fica restrito ao uso de banco de dados estáticos.

Uma indicação para trabalhos futuros seria a combinação da análise temporal das variáveis e execução dos modelos junto a um sistema capaz de gerar dados em tempo real, para testar o desempenho dessas técnicas e determinar se elas são viáveis durante o monitoramento dos poços.

## Referências

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. *Três campos do pré-sal já respondem por mais da metade da produção brasileira de petróleo e gás natural*. 2019. Disponível em: <http://www.anp.gov.br/noticias/anp-e-p/5466-tres-maiores-campos-ja-respondem-por-mais-da-metade-da-producao-brasileira-de-petroleo-e-gas-natural>.

AKINSETE, O; OSHINGBESAN, A. *Leak Detection in Natural Gas Pipelines Using Intelligent Models*. SPE International Heavy Oil Conference and Exhibition held in Kuwait City, Kuwait. 2019.

ALNAIMAT, F. AND ZIAUDDIN, M. *Wax deposition and prediction in petroleum pipelines*. Journal of Petroleum Science and Engineering 184, 2020.

ALOUHALI, R.; ALJUBRAN, M.; GHARBI, S.; AL-YAMI, A. *Drilling Through Data: Automated Kick Detection Using Data Mining*. SPE International Heavy Oil Conference and Exhibition held in Kuwait City, Kuwait. 2018.

APIO, A.; DAMBROS, J. W. V.; DIEHL, F. C.; FARENZENA, M.; TRIERWEILER, J. O. *PDG Pressure Estimation in Offshore Oil Wells: Extended Kalman Filter vs. Artificial Neural Networks*. IFAC 52-1, 2019.

BALAKIN, B. V.; LO, S.; KOSINSKI, P.; HOFFMAN, A. C. *Modelling agglomeration and deposition of gas hydrate in industrial pipelines with combined CFD-PBM technique*. Chemical Engineering Science 153, 2016.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. Springer, 2006.

BRAMER, M. *Principles of data mining*. 3ª Edição. Springer, 2016.

BRILL, J. P.; MUKHERJEE, H. *Multiphase Flow in Wells*. Society of Petroleum Engineer Inc. Texas, 1999.

CAMPBELL, B. C. & LIU, Y. *A nonlinear flow transition criterion for the onset of slugging in horizontal channels and pipes*. Physics of fluids. 2016. DOI: 10.1063/1.4960084.

CHAUDHARI, P.; ZERPA, L. E.; SUM, A. K. *A correlation to quantify hydrate plugging risk in oil and gas production pipelines based on hydrate transportability parameters*. Journal of Natural Gas Science and Engineering 58, 2018.

CHEN, Y.; WU, W. *Separation of geochemical anomalies from the sample data of unknown distribution population using Gaussian Mixture Model*. Computers and Geosciences 125, 2019.



Comprehensive R Archive Network (CRAN). Package ‘adabag’. 2018. Disponível em: <<https://cran.r-project.org/web/packages/adabag/adabag.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘bnlearn’. 2019. Disponível em: <<https://cran.r-project.org/web/packages/bnlearn/bnlearn.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘caret’. 2020. Disponível em: <<https://cran.r-project.org/web/packages/naivebayes/naivebayes.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘corrplot’. 2017. Disponível em: <<https://cran.r-project.org/web/packages/corrplot/corrplot.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘ggplot2’. 2019. Disponível em: <<https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘janitor’. 2020. Disponível em: <<https://cran.r-project.org/web/packages/janitor/janitor.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘naivebayes’. 2019. Disponível em: <<https://cran.r-project.org/web/packages/naivebayes/naivebayes.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘random Forest’. 2018. Disponível em: <<https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘reshape2’. 2017. Disponível em: <<https://cran.r-project.org/web/packages/reshape2/reshape2.pdf>>.

Comprehensive R Archive Network (CRAN). Package ‘rpart.plot’. 2019. Disponível em: <<https://cran.r-project.org/web/packages/rpart.plot/rpart.plot.pdf>>.

COTO, B.; MARTOS, C.; ESPADA, J. J.; ROBUSTILLO, M. D.; PEÑA, J. L. *Analysis of paraffin precipitation from petroleum mixtures by means of DSC: Interactive procedure considering solid-liquid equilibrium equations*. Fuel 89, 2010.

DA SILVA, M. *Impedance Sensors for Fast Multiphase Flow Measurement and Imaging*. ResearchGate, 2014. Disponível em: <[https://www.researchgate.net/publication/44229478\\_Impedance\\_Sensors\\_for\\_Fast\\_Multiphase\\_Flow\\_Measurement\\_and\\_Imaging](https://www.researchgate.net/publication/44229478_Impedance_Sensors_for_Fast_Multiphase_Flow_Measurement_and_Imaging)>

DIAZ-ROZO, J.; BIELZA, C.; LARRAÑAGA, P. *Machine-tool condition monitoring with Gaussian mixture models-based dynamic probabilistic clustering*. Engineering Applications of Artificial Intelligence 89, 2020.

DUAN, J.; LIU, H.; JIANG, J.; XUE, S.; WU, J.; GONG, J. *Numerical prediction of wax deposition in oil-gas stratified pipe flow*. International Journal of Heat and Mass Transfer 105, 2017.

FARID, D. M.; ZHANG, L.; RAHMAN, C. M.; HOSSAIN, M. A.; STRACHAN, R. *Hybrid decision tree and naive Bayes classifier for multi-class classification tasks*. Expert Systems with Applications 41, 2014.

GHASEMIAN, N.; AKHOONDZADEH, M. *Introducing two Random Forest based methods for cloud detection in remote sensing images*. Advances in Space Research 62, 2018.

GLÉNAT, P.; SAHA, P.; MACPHERSON, C.; YANG, J.; TOHIDI, B. *Field Tests Results of Production Inside Hydrate P&T Zone Using New Hydrate Monitoring Instruments*. 12th Offshore Mediterranean Conference and Exhibition in Ravenna. 2015.

GRINGARTEN, A. C.; VON SCHROETER, T.; ROLFSVAAG, T.; BRUNER, J. *Use of Downhole Permanent Pressure Gauge Data to Diagnose Production Problems in a North Sea Horizontal Well*. SPE Annual Technical Conference and Exhibition held in Denver, Colorado, U.S.A., 5 – 8 October 2003.

GUDMUNDSSON, J. S. *Flow Assurance Solids in Oil and Gas Production: The big five*. CRC Press. 2018.

GUGGARI, S.; KADAPPA, V.; UMADEVI, V. *Non-sequential partitioning approaches to decision tree classifier*. Future Computing and Informatics Journal 3, 2018.

GUO, B.; SONG, S.; GHALAMBOR, A.; LIN, T. *Offshore Pipelines*. Gulf Professional Publishing. 2ª Edição, 2013.

GUO, B.; LIU, X.; TAN, X. *Petroleum Production Engineering*. Gulf Professional Publishing, 2017.

HERBRICH, R. *Learning kernel classifiers: theory and algorithms*. MIT Press, 2002.

HOLLAND, F.; BRAGG, R. *Fluid Flow for Chemical Engineers*. 2ª Edição Butterworth-Heinemann, 1995.

HOSSIN, M.; SULAIMAN, M. N. *A review on evaluation metrics for data classification evaluations*. International Journal of Data Mining & Knowledge Management Process (IJDMP) vol. 5, nº2, 2015.

International Energy Agency. *Offshore Energy Outlook 2018*. 2018. IEA, Paris. Disponível em: <https://www.iea.org/reports/offshore-energy-outlook-2018>.

JANSEN, F. E.; SHOHAM, O.; TAITEL, Y. *The elimination of severe slugging: experiments and modeling*. International Journal of Multiphase Flow v. 22, 1996.

WITTEN, I. H.; FRANK, E. *Data mining: A practical machine learning tools and techniques*. 2ª Edição. Elsevier, 2005.

KARAMI, M.; WANG, L. *Fault detection and diagnosis for nonlinear systems: A new adaptative Gaussian mixture modeling approach*. Energy and Buildings 166, 2018.

KHANNA, A.; BURLA, R.; PATWARDHAN, S. *Empirical Correlations for Quick and Accurate Hydrate Prediction – which on to apply?.* SPE Oil and Gas Indian Conference and Exhibition, Mumbai-Índia, 2019.

KHODNENKO, I.; IVANOV, S.; PERETS, D.; SIMONOV, M. *Detection of lost circulation in drilling wells employing sensor data using machine learning technique*. Procedia Computer Science 156. 2019.

LACHANCE, J. W.; KEINATH, B. L. *Hydrate Cold Restarts: Paradigm Shifts in Hydrate Management*. IPTC – International Petroleum Technology Conference, 2015.

LAIK, S. *Offshore Petroleum Drilling and Production*. CRC Press, 2018.

LAZRI, M.; AMEUR, S. *Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVERI data*. Atmospheric Research 203, 2018.

LIMA, C. B. D. C. *Elementos estruturantes das operações integradas na construção de poços de petróleo offshore no ambiente 4.0*. 2018. Dissertação – Escola de Engenharia, Universidade Federal Fluminense, Niterói.

LIU, Y.; GE, Z. *Weighted random forests for fault classification in industrial process with hierarchical clustering model selection*. Journal of Process Control 64, 2018.

LUNA-ORTIZ, E.; LAWRENCE, P.; PANTELIDES, C. C.; ADJIMAN, C. S.; IMMANUEL, C. D. *An integrated framework for model-based flow assurance in deep-water oil and gas production*. ESCAPE 18, 2008.

MAJKA, M. *Package “naivebayes”*. CRAN, 2020. Disponível em: <<https://cran.r-project.org/web/packages/naivebayes/naivebayes.pdf>>.

MARINS, M. A. *Machine Learning Techniques Applied to Hydrate Failure Detection on Production Lines*. Dissertação – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2018.

MARQUES, F. S. B. *Classificação Dinâmica para o Evento de Prisão de Coluna na Perfuração de Poços Offshore*. 2019. Dissertação – Escola de Engenharia, Universidade Federal Fluminense, Niterói.

MARTLOFF, N. *Statistical regression and classification from linear models to machine learning classification*. CRC Press, 2017.

NIELSEN, D. *Tree boosting with XGBoost: Why does XGBoost win “every” machine learning competition?*. 2016. Dissertação – Departamento de Ciências Matemáticas, Norwegian University of Science and Technology.

OLAJIRE, A. A. *A Review of oilfield scale management technology for oil and gas production*. Journal of Petroleum Science and Engineering 135, 2015.

OSOKOGWU, U.; AJIENKA, J. A. *Modeling of Hydrate Dissociation in Subsea Natural Gas Production Flowlines*. SPE Oil International Conference and Exhibition, Tinapa-Nigéria, 2010.

PATEL, E.; KUSHWAHA, D. S. *Clustering cloud workloads: K-Means vs. Gaussian Mixture Model*. Procedia Computer Science 171, 2020.

PEDERSEN, S.; DURDEVIC, P.; YANG, Z. *Challenges in slug modeling and control for offshore oil and gas production: A review study*. International Journal of Multiphase Flow v. 88, 2017.

PEREIRA, E. J. R. *Classificação do Evento de Prisão de Coluna Mecânica em Atividade de Perfuração de Poços Offshore por Técnica de Aprendizado de Máquina*. 2019. Dissertação – Escola de Engenharia, Universidade Federal Fluminense, Niterói.

PETROBRAS. Comparativo entre os diferentes tipos de plataformas. 2014. Disponível em: <https://petrobras.com.br/infograficos/tipos-de-plataformas/desktop/index.html#>

PINTO, V. S. *Desenvolvimento de Solução CRISP-DM para Classificação do Evento Prisão de Coluna no Processo de Perfuração de Poços Offshore*. 2018. Dissertação – Escola de Engenharia, Universidade Federal Fluminense, Niterói.

QIN, H.; SRIVASTAVA, V.; WANG, H.; ZERPA, L. E.; KOH, C. A. *Machine Learning Models to Predict Gas Hydrate Plugging Risks Using Flowloop and Field Data*. Offshore Technology Conference. 2019.

RASCHKA, S. *Naive Bayes and text classification: Introduction and theory*. Cornell University, 2014.

SANTOS, I. H.; LISBOA, H. F.; FEITAL, T. S.; CÂMARA, M. M.; SOARES, R. M.; MARINS, M. A.; BARROS, B. D.; PREGO, T. M.; LIMA, A. A.; NETTO, S. L. *Hydrate Failure Detection in Production and Injection Lines Using Model and Data-driven Approaches*. Rio Oil & Gas Expo and Conference 2018.

SANTOS, M. J. R.; MARQUES, F. S. B.; D’ALMEIDA, A. L.; FERREIRA, G. S.; COSTA, Y. S.; LIMA, G. B. A.; PEREIRA, E. J. R.; LETA, F. R.; LIMA, C. B. C.; PINTO, H. L. C. *Challenges in applying artificial intelligence techniques for stuck pipe*

*detection: a pre-salt case* (Pôster). in: I Seminário de Inteligência Artificial Aplicada à Indústria do Petróleo, 2019.

SCHMIDT, J.; BRILL, J. P.; BEGGS, H. D. *Experimental study of severe slug in two-phase flow pipeline-riser pipe system*. SPE Journal v. 20, issue 05, 1980.

SHI, B.; SONG, S.; LV, X.; LI, W.; WANG, Y.; DING, L.; LIU, Y.; YANG, J.; WU, H.; WANG, W.; GONG, J. *Investigation on natural gas hydrate dissociation from slurry to water-in-oil emulsion in a high pressure flow loop*. Fuel 233, 2018.

SHIPWAY, N. J.; BARDEN, T. J.; HUTHWAITE, P.; LOWE, M. J. S. *Automated defect detection for Fluorescent Penetrant Inspection using Random Forest*. NDT and E International 108, 2019.

SLOAN, D.; KOH, C.; SUM, A. K. *Natural Gas Hydrates in Flow Assurance*. Gulf Professional Publishing. 2011.

SLOAN, E. D.; KOH, C. A.; SUM, A. K.; BALLARD, A. L.; SHOUP, G. J.; MCMULLEN, N.; CREEK, J. L.; PALERMO, T. *Hydrate: state of art inside and outside flowlines*. Journal of Petroleum Technology. Dezembro, 2009.

TAITEL, Y. *Stability of severe slugging*. International Journal of Multiphase Flow, v.12, 1986.

TATTERSALL, S. *Choke valve technology in subsea environments*. Measurement and Control v. 49, 2016.

TIN, V. *Severe slug in flexible riser*. International Conference on Multiphase Production, França, 1991.

THEYAB, M. *Severe Slugging Control: Simulation of a Real Case Study*. Journal of Environmental Research, 2018.

TRABELSI, A.; ELOUEDI, Z.; LEFEVRE, E. *Decision tree classifiers for evidential attribute values and class labels*. Fuzzy Sets and Systems 366, 2019.

VARGAS, R. E. V.; MUNARO, C. J.; CIARELLI, P. M.; ARAÚJO, J. C. D. *Proposal for Two Classifiers of Offshore Naturally Flowing Wells Events Using k-Nearest Neighbors, Sliding Window and Time Multiscale*. 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP) held in Taipei, Taiwan .2017.

VARGAS, R. E. V.; MUNARO, C. J.; CIARELLI, P. M.; MEDEIROS, A. G.; AMARAL, B. G.; BURRIONUEVO, D. C.; ARAÚJO, J. C. D.; RIBEIRO, J. L.; MAGALHÃES, L. P. *A realistic and public dataset with rare undesirable real events in oil wells*. Journal of Petroleum Science and Engineering, v. 181, 2019.

XU, Q.; ZHOU, H.; ZHU, Y.; CAO, Y.; HUANG, B.; LI, W.; GUO, L. *Study of identification of global flow regime in a long pipeline transportation system*. Powder Technology v. 362, 2019.

XUE, W.; JIANG, T. *An adaptative algorithm for target recognition using Gaussian mixture models*. Measurement 124, 2018.

YE, J.; GUO, L. *Multiphase flow pattern recognition in pipeline-riser systems by statistical feature clustering of pressure fluctuation*. Chemistry Engineer Society v. 102, 2013.

YOO, C.; HAN, D.; IM, J.; BECHTEL, B. *Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images*. ISPRS Journal of Photogrammetry and Remote Sensing 157, 2019.

ZHANG, H. *The optimality of Naive Bayes*. American Association for Artificial Intelligence, 2004.

ZOU, S.; GUO, L.; XIE, C. *Fast recognition of global flow regime in pipeline-riser systems by spatial correlation of differential pressures*. Multiphase Flow v. 88, 2017.

PINTO, V. S. Desenvolvimento de Solução CRISP-DM para Classificação do Evento Prisão de Coluna no Processo de Perfuração de Poços Offshore. Dissertação de Mestrado – Escola de Engenharia, Universidade Federal Fluminense, Niterói, 2018.

OILFIELD GLOSSARY. *The Oilfield Glossary: Where the Oilfield Meets the Dictionary*, 1998. Disponível em:< <https://www.glossary.oilfield.slb.com>> Acesso em: 19/10/2020.

PANORAMA OFFSHORE. Riser é imprescindível no mundo offshore, 2019. Disponível em:<<https://www.panoramaoffshore.com.br/riser-e-imprescindivel-no-mundo-offshore/#comments>>. Acesso em: 19/10/2020.

MISHRA, A. *Metrics to Evaluate your Machine Learning Algorithm*. Disponível em:<<https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>> Acesso em