

ANÁLISE DA QUALIDADE DA CIMENTAÇÃO DE POÇOS DE PETRÓLEO OFFSHORE UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Lorena Bicalho Lima

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Civil, COPPE, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Mestre em Engenharia Civil.

Orientadores: Alexandre Gonçalves Evsukoff
Romildo Dias Toledo Filho

Rio de Janeiro
Setembro de 2018

ANÁLISE DA QUALIDADE DA CIMENTAÇÃO DE POÇOS DE PETRÓLEO
OFFSHORE UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Lorena Bicalho Lima

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA CIVIL.

Examinada por:

Prof. Alexandre Gonçalves Evsukoff, D.Sc.

Prof. Romildo Dias Toledo Filho, D.Sc.

Prof. Rogério Pinto Espíndola, D.Sc.

Dra. Cristiane Richard de Miranda, D.Sc.

RIO DE JANEIRO, RJ BRASIL

SETEMBRO DE 2018

Lima, Lorena Bicalho

Análise da Qualidade da Cimentação de Poços de Petróleo Offshore Utilizando Técnicas de Mineração de Dados/ Lorena Bicalho Lima. – Rio de Janeiro: UFRJ/COPPE, 2018.

XXI, 143 p.: il.; 29,7 cm.

Orientadores: Alexandre Gonçalves Evsukoff

Romildo Dias Toledo Filho

Dissertação (mestrado) – UFRJ/ COPPE/ Engenharia Civil, 2018.

Referências Bibliográficas: p. 88-90.

1. Petróleo. 2. Cimentação. 3. Perfilagem. I. Evsukoff, Alexandre Gonçalves *et al.*. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

Agradecimentos

Agradeço ao meu marido Bruno pela paciência nos meus momentos de desespero e por entender a necessidade de abdicação de finais de semana e do convívio social para o desenvolvimento desse trabalho.

Agradeço à minha família pela força e compreensão pelos momentos de ausência.

Agradeço à Petrobras pela oportunidade de realizar esse estudo e pelo acesso aos dados. Ao Thiago Piedade por todo o conhecimento passado ao longo dos anos e, juntamente com o Gilson Campos, o Diego Brasil e o Rafael Peralta pela seleção dos principais parâmetros operacionais que afetam a cimentação, informações de grande importância para o desenvolvimento desse trabalho.

Agradeço aos professores Alexandre e Romildo pela confiança no meu trabalho e dicas imprescindíveis para o desenvolvimento desse trabalho.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ANÁLISE DA QUALIDADE DA CIMENTAÇÃO DE POÇOS DE PETRÓLEO OFFSHORE UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Lorena Bicalho Lima

Setembro/2018

Orientadores: Alexandre Gonçalves Evsukoff

Romildo Dias Toledo Filho

Programa: Engenharia Civil

Este trabalho apresenta o primeiro estudo de correlação entre informações de planejamento e operação da cimentação primária em poços de petróleo *offshore* com o resultado dessas, verificados por meio de perfis acústicos de análise da qualidade da cimentação, utilizando técnicas de mineração de dados. O ganho para a indústria é a confiabilidade em seguir as operações do poço com segurança, sem a necessidade de avaliar a cimentação por meio de perfis, o que gera uma economia significativa. Foi utilizada, primeiramente, uma base de dados construída com as informações dos poços perfilados da Petrobras desde 2013. Duas outras bases de dados foram criadas durante o desenvolvimento do trabalho, com informações que se mostraram mais relevantes para o problema. Este trabalho aponta uma direção para o prosseguimento de estudos dessa natureza, mas indica a necessidade de pesquisas mais aprofundadas para cada cenário, devido à importância dessas informações para a segurança de poço que, se analisadas incorretamente, podem levar a catástrofes.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

CEMENT QUALITY ANALYSE OF OFFSHORE OIL WELL UNDER DATA
MINING TECHNIQUES

Lorena Bicalho Lima

September/2018

Advisors: Alexandre Gonçalves Evsukoff
Romildo Dias Toledo Filho

Department: Civil Engineering

This work presents the first study of correlation between planning and operation information of the primary cement job in offshore oil wells with the result of these, verified through acoustic logs for cement quality analysis. The big advance for the industry is the reliability of safely following the operations of the well, without the need to cement logging evaluation, which generates a significant saving. For this purpose, at first, a database constructed with information from Petrobras' wells since 2013 was used. Two other databases had to be created during the development of this work, with information that showed most relevant to the problem. This work presents a direction for further studies of this nature, but indicates the need for further research for each scenario due to the importance of this information for well safety, which, if analyzed incorrectly, can lead to irreparable catastrophes.

Sumário

1 Introdução	1
1.1 Objetivo	2
1.2 Mineração de Dados	2
2 Entendimento do Negócio	6
2.1 Integridade de Poço	6
2.1.1 Falhas de Integridade de Poço	9
2.2 Cimentação de Poços de Petróleo.....	13
2.2.1 Planejamento da Cimentação	14
2.2.1.1 Informações preliminares	14
2.2.1.2 Aditivos para Cimentação	15
2.2.1.2.1 Acelerador	15
2.2.1.2.2 Retardador	15
2.2.1.2.3 Dispersante	15
2.2.1.2.4 Estendedor	16
2.2.1.2.5 Adensante	16
2.2.1.2.6 Viscosificante	16
2.2.1.2.7 Controlador de Filtrado	16
2.2.1.2.8 Controlador de Gás	16
2.2.1.2.9 Agentes Expansivos.....	17
2.2.1.2.10 Sílica	17
2.2.1.3 Colchões lavadores e espaçadores.....	17
2.2.1.4 Ensaios Laboratoriais	18
2.2.1.4.1 Peso Específico	19
2.2.1.4.2 Filtrado.....	20

2.2.1.4.3 Tempo de Espessamento	20
2.2.1.4.4 Água Livre	20
2.2.1.4.5 Parâmetros Reológicos	20
2.2.1.4.6 Resistência à compressão	21
2.2.2 Operação de Cimentação	21
2.2.3 Fatores que Afetam a Qualidade da Cimentação	22
2.2.3.1 Eficiência de Deslocamento	22
2.2.3.2 Centralização do Revestimento	22
2.2.3.3 Volume da Pasta de Cimento	22
2.2.3.4 Propriedades do fluido de perfuração	23
2.2.3.5 Uso do plugue de topo e de fundo	23
2.2.3.6 Perdas de circulação	24
2.3 Avaliação de qualidade da cimentação através de perfis acústicos	25
2.3.1 Ferramentas Sônicas	25
2.3.1.1 Cement Bond Log (CBL)	25
2.3.1.2 Variable Density Log (VDL)	27
2.3.2 Ferramentas Ultrassônicas	29
2.3.3 Perfis de Cimentação	30
3 O Conjunto de Dados	32
3.1 Análise dos Dados	39
3.1.1 Influência da companhia de cimentação	39
3.1.2 Influência da localização do poço	40
3.1.3 Influência da fase do poço	41
3.1.4 Influência da fase do poço em relação à localização	42
3.1.5 Influência do fluido de perfuração	43

3.1.6	Influência do fluido de perfuração em relação do fabricante do fluido	44
3.1.7	Influência da companhia de cimentação em relação ao fabricante do fluido de perfuração	45
3.1.8	Influência da presença de aditivo químico	46
3.1.9	Influência da inclinação do poço	47
3.1.10	Influência da companhia de cimentação em relação à inclinação do poço .	48
3.1.11	Influência da extensão cimentada.....	49
3.1.12	Influência da ocorrência de perda de circulação	50
3.1.13	Ocorrência de falha na operação de cimentação pela análise dos dados de BD2.....	52
4	Modelagem	54
4.1	Classificação	54
4.1.1	Naive Bayes.....	54
4.1.2	Árvores de Decisão.....	56
4.1.3	Tree Ensemble	60
4.1.4	Redes Neurais	61
4.2	Validação dos Modelos	64
4.2.1	Precisão.....	65
4.2.2	Recuperação	66
4.2.3	Medida F.....	68
4.2.4	AUC.....	68
5	Análise dos Resultados.....	71
5.1	Banco de Dados 1	72
5.2	Banco de Dados 2	76
5.3	Banco de Dados 3	79

5.4 Comparação entre os resultados dos três bancos de dados.....	83
6 Conclusões.....	85
7 Referências bibliográficas	88
APÊNDICE A - Lista de atributos do BD1	91
APÊNDICE B - Lista de atributos do BD2	99
APÊNDICE C - Lista de atributos do BD3	102
APÊNDICE D – Resumo das informações presente em BD1, BD2 e BD3.....	109
APÊNDICE E – Poços com Falha de Isolamento	111
APÊNDICE F – Matrizes de confusão	114

Lista de Figuras

Figura 1 - Fases da metodologia CRISP-DM (Traduzido de [9])	3
Figura 2 - Tarefas de mineração de dados [15]	5
Figura 3 - Possíveis caminhos de fluxo [10]	7
Figura 4 - Exemplos de atalhos de vazamentos contornado barreiras de segurança individuais. [10].....	7
Figura 5 - Perda de integridade por categoria de falha (Traduzido de [28])	10
Figura 6 - Ocorrências de blowouts na OCS [19]	11
Figura 7 - Compilação de trabalhos que indicam perda de integridade por categoria de falha (Traduzido de [3])	11
Figura 8 - Fases de monitoramento da integridade do poço (Traduzido de [3])	12
Figura 9 - Cimentação de um poço de petróleo (Adaptado de [17])	13
Figura 10 - Fluxograma dos principais testes de pasta [27]	19
Figura 11 - Trajetória de energia acústica entre transmissor e receptor a 3 pés [11]	26
Figura 12 - Onda composta recebida no transdutor a 3 pés [11]	26
Figura 13 - Trajetória de energia acústica entre transmissor e receptor a 5 pés [11]	27
Figura 14 – Microsismograma VDL (Adaptado de [4]).....	28
Figura 15 - Princípio de funcionamento das ferramentas ultrassônicas	29
Figura 16 - Perfil de Avaliação da Qualidade da Cimentação [1].....	31
Figura 17 – Classificação da qualidade da cimentação BD1 – 5 classes	33
Figura 18 – Classificação da qualidade da cimentação BD1 – 4 classes	34
Figura 19 – Classificação da qualidade da cimentação BD1 – 2 classes	34
Figura 20 – Classificação da qualidade da cimentação BD2 – 4 classes	37

Figura 21 – Classificação da qualidade da cimentação BD2 – 2 classes	37
Figura 22 – Classificação da qualidade da cimentação BD3 – 4 classes	38
Figura 23 – Classificação da qualidade da cimentação BD3 – 2 classes	38
Figura 24 – Análise da qualidade da cimentação por companhia de serviço	40
Figura 25 – Análise da qualidade da cimentação por localização do poço	41
Figura 26 – Análise da qualidade da cimentação por fase do poço.....	42
Figura 27 – Análise da qualidade da cimentação por cenário considerando localização e fase do poço	43
Figura 28 – Análise da qualidade da cimentação por tipo de fluido de perfuração	44
Figura 29 – Análise da qualidade da cimentação por cenário considerando tipo e fabricante do fluido de perfuração.....	45
Figura 30 – Análise da qualidade da cimentação por cenário considerando fabricante do fluido de perfuração e companhia de cimentação	46
Figura 31 – Análise da qualidade da cimentação por presença de aditivo químico na pasta de cimento.....	47
Figura 32 – Análise da qualidade da cimentação por faixa de inclinação do poço	48
Figura 33 – Análise da qualidade da cimentação por cenário considerando faixa de inclinação do poço e companhia de cimentação	49
Figura 34 – Análise da qualidade da cimentação por extensão cimentada	50
Figura 35 - Relação entre perda de circulação, volume de excesso de pasta bombeado e a obtenção de isolamento hidráulico em BD1.....	51
Figura 36 - Recorrência de falhas em BD2 para poços com falha de isolamento hidráulico	53
Figura 37 - Workflow Naive-Bayes	55
Figura 38 – Configuração do modelo de Naive Bayes	56

Figura 39 – Representação da árvore de decisão e particionamento do espaço de atributos [14]	57
Figura 40 – Workflow Árvore de Decisão	57
Figura 41 – Configuração do modelo de Árvore de Decisão	59
Figura 42 – Workflow Tree Ensemble	61
Figura 43 – Configuração do modelo Tree Ensemble	61
Figura 44 – Funcionamento de um neurônio artificial (Traduzido de [16]).....	62
Figura 45 – Exemplo de rede MLP [16].....	63
Figura 46 – Workflow Redes Neurais	64
Figura 47 – Configuração do modelo de Redes Neurais	64
Figura 48 - Espaço ROC.....	68
Figura 49 – AUC	69

Lista de Tabelas

Tabela 1 - Impedância acústica características dos materiais (Traduzido de [13])	30
Tabela 2 – Classificação da qualidade da cimentação.....	32
Tabela 3 – Lista de valores ausentes em BD1	35
Tabela 4 - Exemplo de Matriz de Confusão (Adaptado de [12])	65
Tabela 5 – Resumo da quantidade de modelos de classificação testados	72
Tabela 6 – Comparação de desempenho de classificação de BD1 com 2 classes com diferentes configurações do modelo de Árvore de Decisão	73
Tabela 7 - Comparação de desempenho de classificação de BD1 com 4 classes com diferentes configurações do modelo de Árvore de Decisão	73
Tabela 8 - Desempenho da classificação de BD1 com 2 classes por Redes Neurais.....	74
Tabela 9 - Desempenho da classificação de BD1 com 4 classes por Redes Neurais.....	75
Tabela 10 - Desempenho da classificação de BD1.....	76
Tabela 11 – Comparação de desempenho de classificação de BD2 com 2 classes com diferentes configurações do modelo de Árvore de Decisão	77
Tabela 12 - Comparação de desempenho de classificação de BD2 com 4 classes com diferentes configurações do modelo de Árvore de Decisão	77
Tabela 13 - Desempenho da classificação de BD2 com 2 classes por Redes Neurais.....	78
Tabela 14 - Desempenho da classificação de BD2 com 4 classes por Redes Neurais.....	78
Tabela 15 - Desempenho da classificação de BD2.....	79
Tabela 16 – Comparação de desempenho de classificação de BD3 com 2 classes com diferentes configurações do modelo de Árvore de Decisão	80

Tabela 17 - Comparação de desempenho de classificação de BD3 com 4 classes com diferentes configurações do modelo de Árvore de Decisão	80
Tabela 18 - Desempenho da classificação de BD3 com 2 classes por Redes Neurais.....	81
Tabela 19 - Desempenho da classificação de BD3 com 4 classes por Redes Neurais.....	82
Tabela 20 - Desempenho da classificação de BD3.....	82
Tabela 21 - Desempenho da classificação comparativo para os três bancos de dados utilizando o modelo Naive Bayes.....	84
Tabela 22 - Matriz de confusão para Naive Bayes em BD1 com 2 classes	114
Tabela 23 - Matriz de Confusão para Naive Bayes em BD1 com 4 classes.....	114
Tabela 24 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, sem poda, ponto médio de particionamento.....	115
Tabela 25 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, sem poda, máximo ganho de particionamento	115
Tabela 26 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, com poda, ponto médio de particionamento	115
Tabela 27 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, com poda, máximo ganho de particionamento.....	116
Tabela 28 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, sem poda, ponto médio de particionamento.....	116
Tabela 29 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, sem poda, máximo ganho de particionamento.....	116
Tabela 30 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, com poda, ponto médio de particionamento	117
Tabela 31 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, com poda, máximo ganho de particionamento	117

Tabela 32 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gini Index, sem poda, ponto médio de particionamento	117
Tabela 33 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gini Index, sem poda, máximo ganho de particionamento	118
Tabela 34 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gini Index, com poda, ponto médio de particionamento	118
Tabela 35 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gini Index, com poda, máximo ganho de particionamento.....	119
Tabela 36 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gain Ratio, sem poda, ponto médio de particionamento.....	119
Tabela 37 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gain Ratio, sem poda, máximo ganho de particionamento.....	120
Tabela 38 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gain Ratio, com poda, ponto médio de particionamento	120
Tabela 39 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes	
– Gain Ratio com poda, máximo ganho de particionamento	121
Tabela 40 - Matriz de Confusão para Tree Ensemble em BD1 com 2 classes.	121
Tabela 41 - Matriz de Confusão para Tree Ensemble em BD1 com 4 classes.	122
Tabela 42 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD1 com 2 classes.....	122
Tabela 43 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD1 com 2 classes.....	122
Tabela 44 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD1 com 2 classes	123
Tabela 45 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD1 com 2 classes	123
Tabela 46 - Matriz de Confusão para o Redes Neurais (1 camada com 10 neurônios) em BD1 com 4 classes.....	123

Tabela 47 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD1 com 4 classes.....	124
Tabela 48 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD1 com 4 classes	124
Tabela 49 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD1 com 4 classes	125
Tabela 50 - Matriz de confusão para Naive Bayes em BD2 com 2 classes	125
Tabela 51 - Matriz de Confusão para Naive Bayes em BD2 com 4 classes.....	126
Tabela 52 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gini Index, sem poda.....	126
Tabela 53 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gini Index, com poda	126
Tabela 54 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gain Ratio, sem poda	127
Tabela 55 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gain Ratio, com poda	127
Tabela 56 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes– Gini Index, sem poda.....	127
Tabela 57 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes – Gini Index, com poda	128
Tabela 58 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes – Gain Ratio, sem poda	128
Tabela 59 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes – Gain Ratio, com poda	129
Tabela 60 - Matriz de Confusão para Tree Ensemble em BD2 com 2 classes.	129
Tabela 61 - Matriz de Confusão para Tree Ensemble em BD2 com 4 classes.	129
Tabela 62 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD2 com 2 classes.....	130

Tabela 63 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD2 com 2 classes.....	130
Tabela 64 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD2 com 2 classes	130
Tabela 65 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD2 com 2 classes	131
Tabela 66 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD2 com 4 classes.....	131
Tabela 67 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD2 com 4 classes.....	131
Tabela 68 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD2 com 4 classes	132
Tabela 69 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD2 com 4 classes	132
Tabela 70 - Matriz de confusão para Naive Bayes em BD3 com 2 classes	133
Tabela 71 - Matriz de Confusão para Naive Bayes em BD3 com 4 classes.....	133
Tabela 72 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes– Gini Index, sem poda, ponto médio de particionamento.....	134
Tabela 73 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gini Index, sem poda, máximo ganho de particionamento	134
Tabela 74 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gini Index, com poda, ponto médio de particionamento	134
Tabela 75 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gini Index, com poda, máximo ganho de particionamento.....	135
Tabela 76 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gain Ratio, sem poda, ponto médio de particionamento.....	135
Tabela 77 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gain Ratio, sem poda, máximo ganho de particionamento.....	135

Tabela 78 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes	
– Gain Ratio, com poda, ponto médio de particionamento	136
Tabela 79 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes	
– Gain Ratio, com poda, máximo ganho de particionamento	136
Tabela 80 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gini Index, sem poda, ponto médio de particionamento	136
Tabela 81 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gini Index, sem poda, máximo ganho de particionamento	137
Tabela 82 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gini Index, com poda, ponto médio de particionamento	137
Tabela 83 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gini Index, com poda, máximo ganho de particionamento	138
Tabela 84 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gain Ratio, sem poda, ponto médio de particionamento	138
Tabela 85 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gain Ratio, sem poda, máximo ganho de particionamento	139
Tabela 86 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gain Ratio, com poda, ponto médio de particionamento	139
Tabela 87 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes	
– Gain Ratio com poda, máximo ganho de particionamento	140
Tabela 88 - Matriz de Confusão para Tree Ensemble em BD3 com 2 classes.	140
Tabela 89 - Matriz de Confusão para Tree Ensemble em BD3 com 4 classes.	140
Tabela 90 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD3 com 2 classes.....	141
Tabela 91 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD3 com 2 classes.....	141
Tabela 92 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD3 com 2 classes	141

Tabela 93 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD3 com 2 classes 142

Tabela 94 - Matriz de Confusão para o modelo de Redes Neurais (1 camada com 10 neurônios) em BD3 com 4 classes..... 142

Tabela 95 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD3 com 4 classes..... 142

Tabela 96 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD3 com 4 classes 143

Tabela 97 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD3 com 4 classes 143

Lista de abreviaturas e siglas

FAPERJ – Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

ANP – Agência Nacional do Petróleo

SGIP – Sistema de Gerenciamento de Integridade de Poços

CSB – Conjunto Solidário de Barreiras

1 Introdução

A preocupação com a segurança de poços é algo cada vez mais forte na indústria do Petróleo. Além da conscientização das operadoras quanto à importância do tema, os órgãos regulamentadores, brasileiros e internacionais, cobram, a cada ano que passa, maior responsabilidade das companhias exploradoras de óleo e gás.

O Acidente de Macondo, ocorrido em abril de 2010 é reconhecido internacionalmente como um dos maiores acidentes da área. Investigações apresentaram resultados que indicam que um maior cuidado com a análise das barreiras de segurança dos poços pode evitar eventos catastróficos. As análises do Acidente de Macondo levantaram potenciais causas para o ocorrido. Uma dessas causas, descrita detalhadamente no item 5A do Relatório de Investigação do Acidente do “Deepwater Horizon” emitido pela BP [7], está relacionada à falta de integridade do poço pela não obtenção ou falha de isolamento hidráulico do cimento.

A atual situação vivida no mundo do petróleo, com o preço do barril muito abaixo do vivenciado há alguns anos, impõe à indústria a necessidade de otimização dos recursos, de forma a balancear as operações garantindo a segurança e a viabilidade econômica do poço.

A economia conquistada pela supressão da operação de perfilagem para avaliar a qualidade da cimentação dependerá da extensão a ser perfilada e da profundidade máxima a ser atingida, fatores que influenciarão no custo do perfil, a ser pago às companhias de serviço responsáveis pela perfilagem, e no tempo da perfilagem, que está relacionado ao custo diário das sondas de perfuração. Considerando apenas o custo médio das sondas de perfuração semissubmersíveis, divulgados em [18], a economia por operação de perfilagem está na ordem de 94.000 dólares, para operações com uma média de duração de 15 horas. A esses valores ainda devem ser somados os custos diretos relacionados à operação de perfilagem.

1.1 Objetivo

Este trabalho tem como objetivo buscar uma relação entre os parâmetros da cimentação, desde o planejamento até a operação, de poços de petróleo com o resultado obtido através da inferência de isolamento hidráulico através de perfis acústicos.

A partir da correlação entre as variáveis e utilizando técnicas de mineração de dados, desenvolve-se um estudo de modelos de previsão que possam indicar se é necessária a perfilagem de um determinado poço ou se é possível inferir o isolamento hidráulico pelo cimento apenas pela análise de dados de planejamento e da operação da cimentação, de forma a garantir a segurança, sem onerar o custo do poço desnecessariamente.

As análises desenvolvidas neste trabalho utilizaram as informações das operações de cimentação da Petrobras. No entanto, todas as opiniões e conclusões aqui apresentadas são pessoais e, portanto, de responsabilidade do autor e não representam o posicionamento da Companhia.

1.2 Mineração de Dados

A Mineração de Dados consiste na análise de grandes conjuntos de dados de forma a buscar correlação entre eles, identificando padrões e gerando modelos que representem a realidade do problema estudado, transformando dessa forma, dados em conhecimento.

Uma das metodologias desenvolvidas para exploração de dados é o CRISP-DM (*Cross Industry Standard Process for Data Mining*) [9], utilizada neste trabalho. Essa metodologia consiste em seis passos principais, representados pela Figura 1. O modelo CRISP-DM trata os problemas de mineração de dados como cíclicos, não sendo, no entanto, a sequência entre as fases rigorosas, permitindo ir e voltar a uma determinada fase diversas vezes, conforme a necessidade de cada projeto.

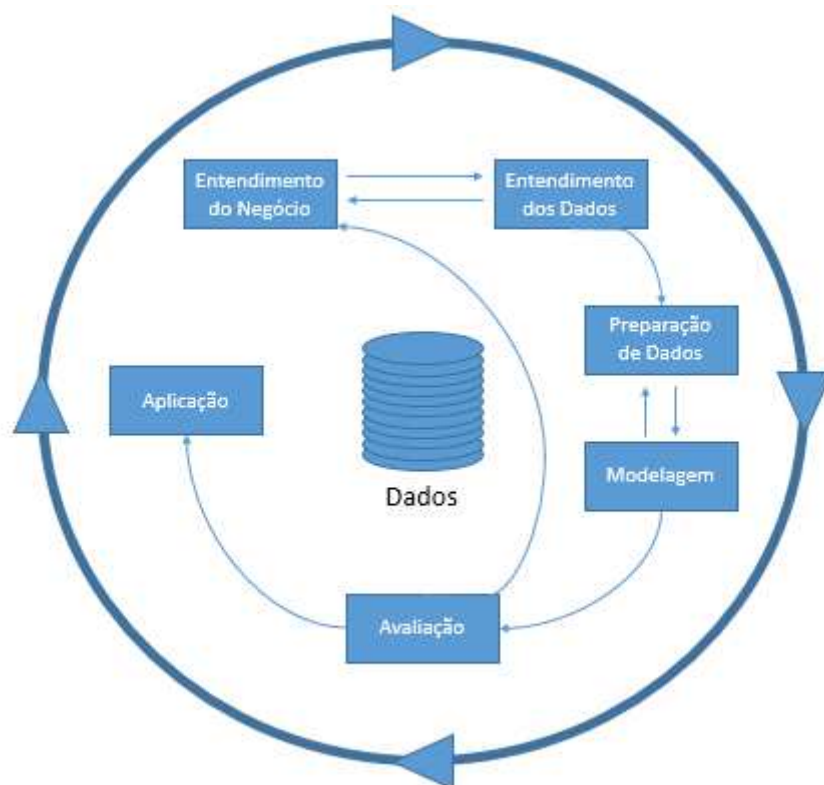


Figura 1 - Fases da metodologia CRISP-DM (Traduzido de [9])

- 1- A primeira etapa é o Entendimento do Negócio, ou seja, a compreensão do problema a ser resolvido e quais os objetivos propostos.
- 2- A partir desse entendimento, é necessário compreender os dados disponíveis, quais as variáveis e como elas representam o problema. Coletar, descrever – usando estatísticas – e verificar a qualidade dos dados.
- 3- Conhecendo os dados, a base deve ser preparada para a modelagem. Essa etapa consiste na seleção dos dados, tratamento do formato, manipulação das variáveis de forma a melhor representarem o problema estudado e, caso necessário, a integração de dados presentes em diferentes bases.
- 4- A quarta etapa é quando os algoritmos serão escolhidos e os modelos serão construídos.
- 5- A etapa de avaliação consiste em analisar os resultados obtidos pelo modelo. Caso tenham sido construídos vários modelos, seus resultados deverão ser

comparados. Os objetivos propostos na primeira etapa podem não ter sido alcançados e, nesse caso, será necessário entender o que pode ter levado à falha e verificar um novo caminho a ser seguido.

- 6- Caso a avaliação dos modelos indiquem o sucesso na obtenção dos objetivos, o modelo pode ser aplicado. Os resultados devem continuar sendo sempre monitorados.

Os modelos de dados, construídos na etapa 5 da metodologia CRISP-DM, e que serão a representação matemática da situação real, são obtidos por algoritmos que podem ser rotulados pelo tipo de tarefa para o qual foi desenvolvido. A primeira etapa de rotulação é a separação entre problemas preditivos, em que existem variáveis de saída que podem ser preditas pelo modelo, e problemas descritivos, em que não existem variáveis de saída. [12]

Entre os problemas preditivos, podemos dividir em classificação e regressão. Na classificação, a variável de saída, geralmente nominal, representa uma determinada categoria. O objetivo então é que modelo consiga prever qual a categoria que um determinado exemplar das variáveis de entrada representa. O problema de regressão também é um problema preditivo, mas, nesse caso, as variáveis de saída são, geralmente, numéricas.

As atividades descritivas podem ser: regras de associação, para as quais deseja-se encontrar relacionamento entre os dados; análises de agrupamento, sem categorias previamente definidas. Nesse caso, os próprios atributos irão definir a qual grupo aquele registro pertence; e sumarização, processo de totalização de registros do conjunto de dados.

A Figura 2 apresenta um resumo da classificação das tarefas de mineração de dados descritas.

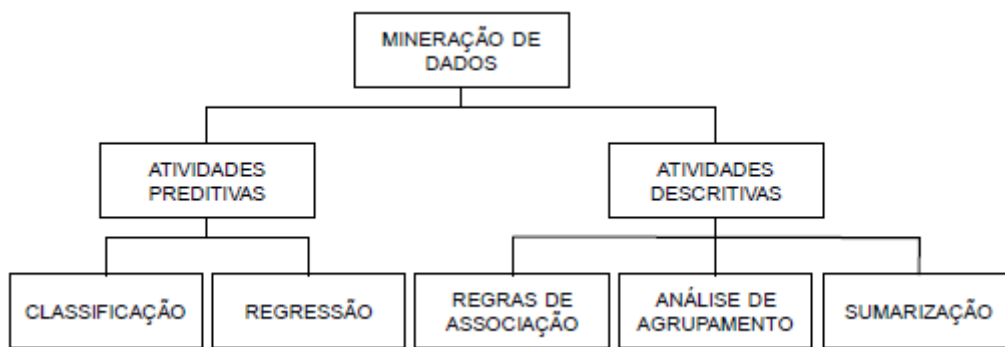


Figura 2 - Tarefas de mineração de dados [15]

2 Entendimento do Negócio

2.1 Integridade de Poço

A Integridade de Poço é definida pela Norma Norueguesa NORSOK D-010 [23] como a aplicação de soluções técnicas, operacionais e organizacionais, para reduzir o risco de descargas não controladas de fluidos do reservatório ao longo do ciclo da vida do poço. Ou seja, objetiva-se prevenir o fluxo indesejado de fluidos entre formações ou para a superfície, de forma a evitar acidentes que possam gerar danos aos trabalhadores e ao meio ambiente.

Um Conjunto Solidário de Barreiras (CSB) pode ser definido como “ um conjunto de um ou mais elementos com o objetivo de impedir o fluxo não intencional de fluidos da formação para o meio externo e entre intervalos no poço, considerando todos os caminhos possíveis” [2], garantindo assim a integridade do poço.

Os possíveis caminhos de fluxo indesejado identificados são enumerados a seguir e ilustrados na Figura 3:

- 1- Interior da coluna de produção, de injeção, de perfuração ou de trabalho
- 2- Interior do poço ou anular de uma das colunas acima
- 3- Anulares dos revestimentos já instalados
- 4- Através das próprias formações

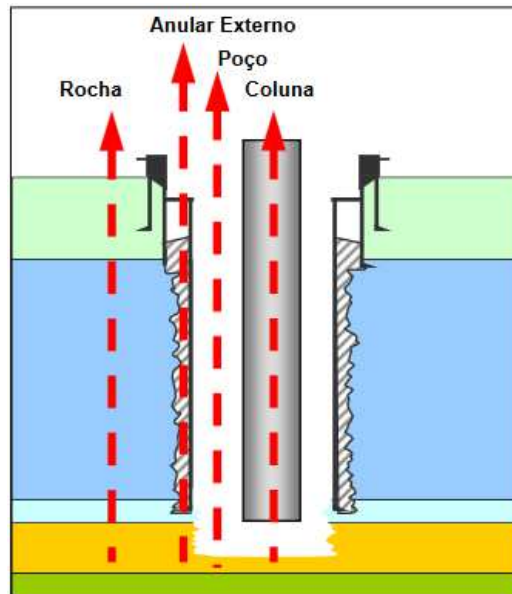


Figura 3 - Possíveis caminhos de fluxo [10]

A possibilidade de “atalhos” entre os caminhos é levantada por Fonseca [10], caso as barreiras existentes em cada caminho não tenham interligação com as barreiras presentes nos demais caminhos possíveis. São possíveis “atalhos” diversos, e dois deles são ilustrados na Figura 4. Portanto, daí advém a necessidade de se falar em “Conjunto Solidário de Barreiras”, em que um conjunto de elementos é combinado de forma a compor uma barreira de segurança e não apenas em barreiras individuais de segurança.

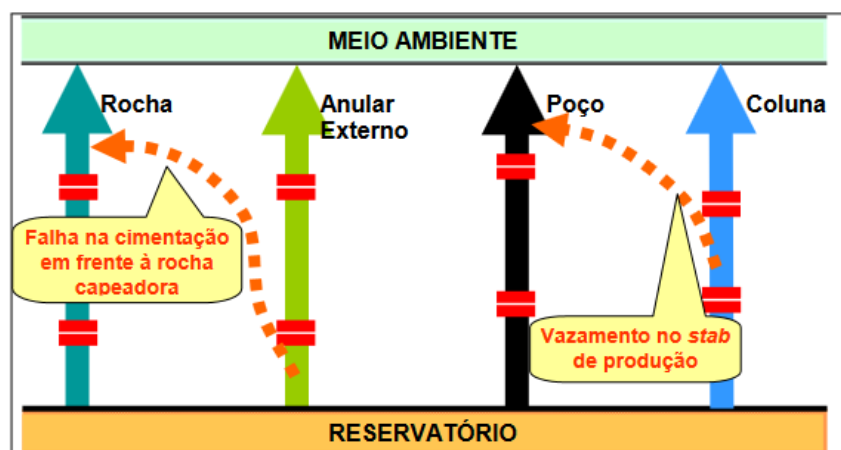


Figura 4 - Exemplos de atalhos de vazamentos contornado barreiras de segurança individuais. [10]

O Regulamento Técnico do Sistema de Gerenciamento da Integridade de Poços (SGIP), oficializado por meio da Resolução Nº 46 de 1º de novembro de 2016 pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) [2], visa a garantia da integridade dos poços mediante o estabelecimento de requisitos e diretrizes de segurança a fim de proteger a vida humana, o meio ambiente e o patrimônio. Dentre os requisitos definidos, se destacam para as análises apresentadas neste trabalho, os critérios para a construção dos CSBs. O SGIP deve ser aplicado durante todo o ciclo de vida dos poços destinados às atividades de Exploração e Produção (E&P) de petróleo e gás natural, desde o seu projeto até o abandono.

O item 10.5 do SGIP, que trata sobre a etapa de abandono, tem como objetivo garantir o isolamento dos intervalos que apresentem Potencial de Fluxo, atual e futuro. São diferenciados os tipos de abandono em permanente e temporário, segundo as seguintes definições presentes no regulamento [2]:

Abandono Permanente de Poços: Situação de um poço na qual há estabelecimento dos Conjuntos Solidários de Barreiras Permanentes e não existe interesse de reentrada futura.

Abandono Temporário de Poços: Situação de um poço na qual há o estabelecimento dos Conjuntos Solidários de Barreiras Temporárias. Adicionalmente, são considerados abandonados temporariamente, poços produtores e injetores já equipados que estejam aguardando o início da produção ou injeção e os poços já em produção que, por algum motivo, encontram-se fechados.

Para o abandono permanente de poços, o SGIP requer que as formações com potencial de fluxo conectadas pela perfuração do poço devem ser isoladas estabelecendo no mínimo 1 CSB entre elas para impedir o fluxo cruzado de fluidos de formações não conectadas naturalmente. No caso de haver formação com potencial de fluxo capaz de fraturar alguma formação acima dela, deve-se estabelecer no mínimo 2 CSBs. Para impedir o fluxo de fluido de qualquer formação para o meio externo deve se estabelecer também 2 CSBs.

Este trabalho tem como foco a análise do elemento de CSB que compõe a barreira no anular entre revestimento-formação ou entre colunas de revestimento. A composição dos materiais utilizados como elemento de CSB devem atender pelo menos

aos seguintes critérios: ser impermeáveis a fluidos, ter propriedade de isolamento que não deteriore com o tempo, ter propriedades mecânicas adequadas para acomodação das cargas a que estarão sujeitos, que não sofram contração que comprometa a integridade e sejam aderentes ao revestimento e formações ao seu entorno. [2] O material utilizado para esse fim, atualmente, é o cimento.

Para garantir a integridade das barreiras de segurança, devem ser definidos, dentre outras coisas, critérios de aceitação e planos de verificação que garantam que os CSBs estão íntegros. A verificação é a confirmação de cada elemento do CSB por meio de avaliação pós instalação ou de observações registradas durante sua instalação. Os processos de verificação podem ser do tipo Teste, em que a verificação do CSB é realizada por aplicação de pressão no sentido do fluxo, considerando pressão diferencial igual ou maior do que a máxima prevista, ou do tipo Confirmação, em que a verificação é realizada através da avaliação dos dados recolhidos durante e/ou após a sua instalação. [2]

Para verificação do cimento como elemento de barreira no anular, existem então as seguintes opções:

- 1- Atestado por normalidade operacional, avaliando os parâmetros de planejamento e operação da cimentação;
- 2- Avaliação através de perfis acústicos de aderência, a partir dos quais é possível inferir a qualidade do cimento no anular, após sua colocação;
- 3- Teste direto de fluxo, para o qual é necessário realizar o canhoneio do revestimento, provocando um ponto de menor resistência na coluna, o que não é desejado e, portanto, é utilizado como última opção.

2.1.1 Falhas de Integridade de Poço

As falhas nos elementos que compõem as barreiras de segurança podem levar à perda de integridade do poço e, conseqüentemente, eventos catastróficos, que podem gerar danos humanos e ao meio ambiente, como ocorrido no Acidente de Macondo, já mencionado nesse trabalho.

Estudo divulgado em 2008 [28] apresenta as principais causas de falhas de integridade de poços offshore na Noruega. Foram analisados 406 poços, dentre os quais 48 produtores e 27 injetores, operados por 7 companhias. Para 18% (75) dos poços analisados, foi reportada a possibilidade de perda de integridade, categorizadas pelo elemento de barreira com falha, conforme Figura 5. Observa-se que a maior ocorrência de perda de integridade é devida à falhas na coluna de produção, seguida de falhas na ASV (válvula de segurança anular), falhas no revestimento e falhas na cimentação.

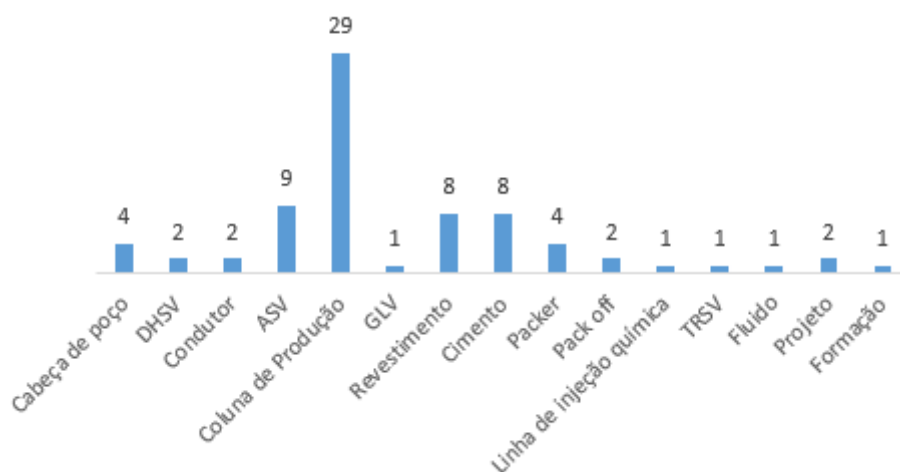


Figura 5 - Perda de integridade por categoria de falha (Traduzido de [28])

Estudo desenvolvido por Izon, em [19] apresenta um resumo das 39 ocorrências de *Blowouts* (vazamento descontrolado de fluido para o meio ambiente) durante operações de perfuração na costa americana de poços do Outer Continental Shelf (OCS) entre 1992 e 2006. Dentre as 39 ocorrências, 18 delas, quase 50%, foram associadas a problemas relacionados à migração de gás durante ou após a cimentação do revestimento, conforme Figura 6.

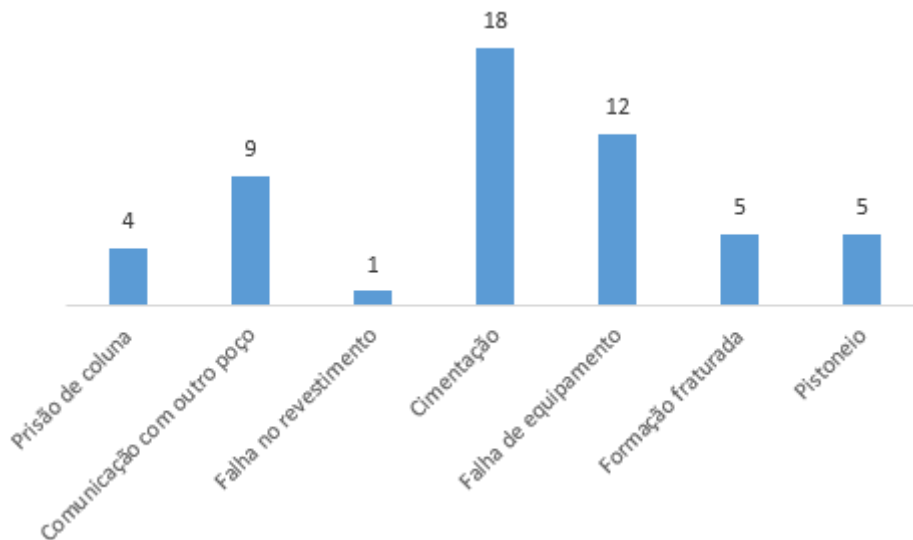


Figura 6 - Ocorrências de blowouts na OCS [19]

AlAwad e Mohammad [3] apresentam uma compilação dos estudos estatísticos de falha de integridade de poços já realizados. A Figura 7 apresenta a distribuição em categorias de elemento de barreira das referências estudadas pelos autores, em que a frequência de ocorrência de falhas relacionadas ao elemento cimento se destaca entre as demais categorias.

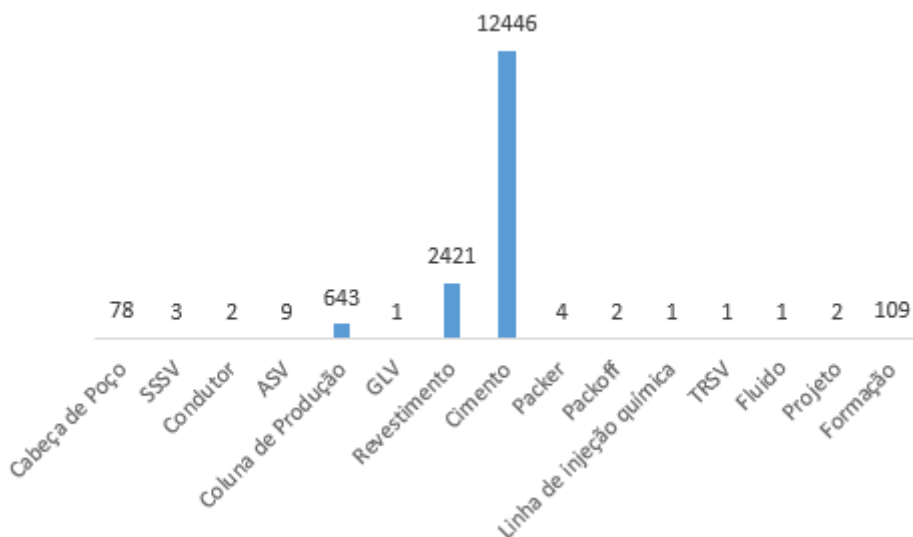


Figura 7 - Compilação de trabalhos que indicam perda de integridade por categoria de falha (Traduzido de [3])

As perdas de integridade do poço pela falha de qualquer elemento de barreira, pode acontecer durante qualquer etapa do poço, e, portanto, sua integridade deve ser monitorada durante todo o ciclo de vida. A Figura 8, apresenta esse monitoramento dividido em quatro etapas: Projeto, Construção, Produção/Injeção e “Workover”, e Abandono do Poço [3]. Este trabalho tem como foco a etapa de construção de poço. Dessa forma, o elemento de CSB identificado durante essa fase, necessita de monitoramento durante o restante do ciclo de vida do poço, já que podem haver indícios de falha devido às condições as quais será exposto.



Figura 8 - Fases de monitoramento da integridade do poço (Traduzido de [3])

Não foram encontrados na literatura estudos de modelagem entre as características de planejamento e operação da cimentação com o resultado da qualidade da cimentação observado em perfil de aderência acústica, conforme propõe esse trabalho.

2.2 Cimentação de Poços de Petróleo

A construção de poços de petróleo conta com etapas de cimentação, as quais visam preencher o espaço anular entre o revestimento e a formação exposta a cada fase do poço. Essas cimentações têm objetivos estruturais e, o mais importante, prover isolamento hidráulico, impedindo o fluxo de fluidos de uma formação para outra ou para superfície e, portanto, podendo constituir um elemento de barreira. Caso os objetivos da cimentação não sejam atendidos, é necessária uma operação para correção desta, a qual é normalmente difícil e muito onerosa. Portanto, o planejamento e a operação da cimentação requerem bastante cuidado e são atividades críticas durante a construção dos poços.

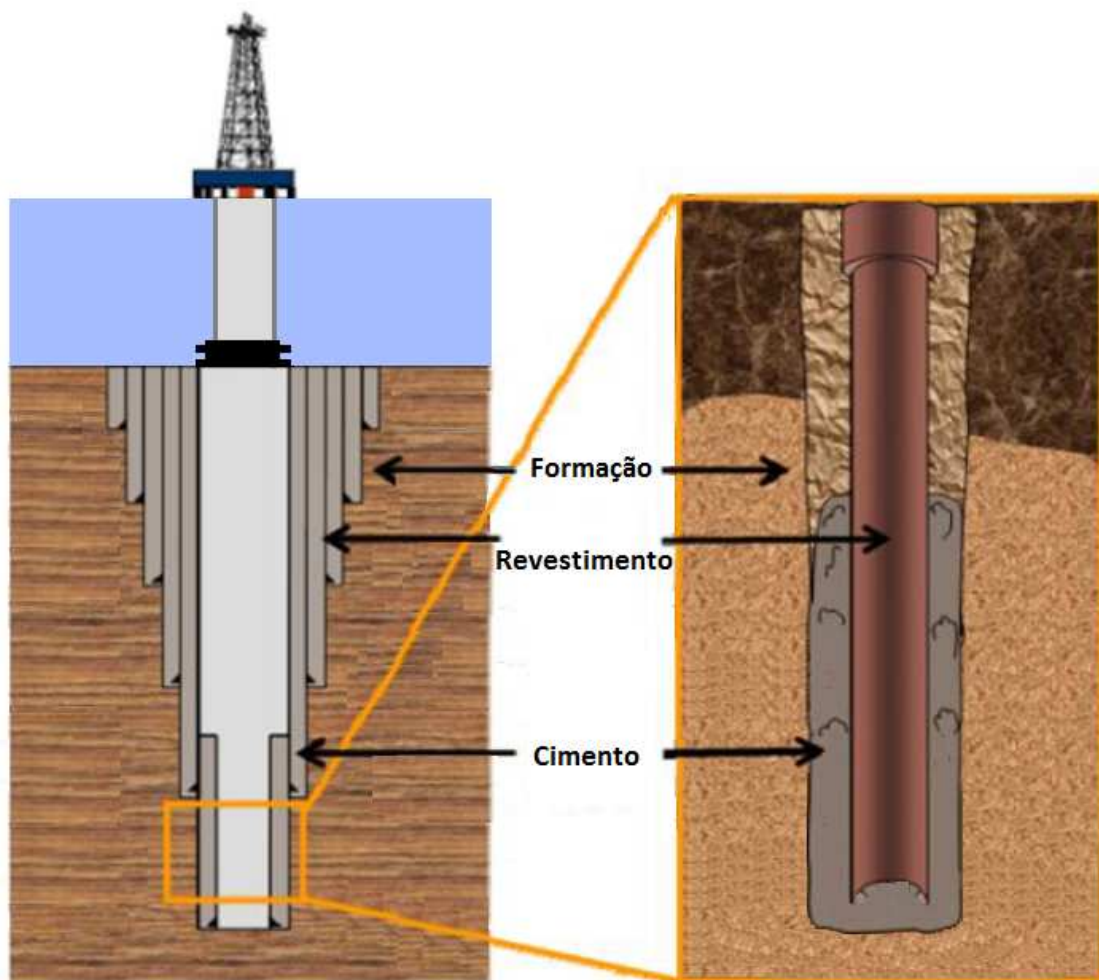


Figura 9 - Cimentação de um poço de petróleo (Adaptado de [17])

2.2.1 Planejamento da Cimentação

O projeto de pastas de cimento deve ser feito para atender às exigências em três situações: sua preparação, deslocamento e pós colocação no anular [5]. Para cada uma dessas etapas, as pastas devem ter valores específicos de suas propriedades e, portanto, antes de qualquer operação de cimentação, é imprescindível a realização de testes laboratoriais e simulações para se conhecer as condições da pasta de cimento a ser bombeada no poço.

2.2.1.1 Informações preliminares

Informações das características do poço, dos fatores aos quais a pasta de cimento será submetida e dos objetivos da cimentação são imprescindíveis para um desenho da pasta de cimento que atenda ao cenário.

A pressão e a temperatura do poço afetam propriedades da pasta de cimento como o tempo de bombeabilidade e o desenvolvimento de resistência compressiva.

As informações de diâmetro do revestimento e diâmetro do poço (“caliper”) irão influenciar o volume de pasta a ser bombeado, que dependerá também do topo de cimento requerido, a fim de cobrir todas as formações permeáveis, provendo o isolamento hidráulico dessas. Em casos em que não é corrido o perfil caliper, pode-se estimar o arrombamento do poço pelas características das formações atravessadas na fase, do fluido no poço, além de outros parâmetros operacionais da perfuração da fase.

Informações de caliper são importantes também para se conhecer a geometria do poço. Um poço que contenha diversos pontos alargados, não irá permitir uma remoção do fluido de perfuração eficiente por se tornarem, nessas regiões, gelificados. Poços espiralados por sua vez prejudicarão a centralização do revestimento, o que também dificulta o deslocamento do fluido de perfuração.

O conhecimento da janela operacional é importante para definição do peso específico da pasta de cimento. A pressão hidrostática da coluna de cimento deve ser inferior à pressão de fratura da formação mais fraca no trecho anular de poço aberto ou ainda menor que a pressão de propagação de fraturas existentes, seja devido aos testes de absorção realizados na sapata do último revestimento ou naturais. Por outro lado, a

pressão hidrostática da coluna de cimento deve ser superior à maior pressão de poros no intervalo, evitando assim uma perda de fluidos para formação.

As propriedades reológicas do fluido de perfuração irão interferir na qualidade da sua remoção e, conseqüentemente, na qualidade da cimentação.

2.2.1.2 Aditivos para Cimentação

As pastas de cimento bombeadas nos poços de petróleo são submetidas a diversas condições, dependendo da configuração de cada um deles. Para que os objetivos da operação sejam alcançados é então necessário a utilização de aditivos, ou seja, produtos químicos específicos que melhoram uma determinada característica da pasta de cimento, enquanto fluida ou após o desenvolvimento da cura.

2.2.1.2.1 Acelerador

Aceleradores são adicionados às pastas de cimento para diminuir o tempo de hidratação da pasta de cimento, ou seja, o tempo que o material demora para se tornar sólido. São utilizados no caso de cimentações de pequena profundidade para que o tempo de espera para a retomada da perfuração não seja muito longo devido ao efeito da baixa temperatura, o qual aumenta o tempo de hidratação da pasta de cimento.

2.2.1.2.2 Retardador

Os retardadores agem em sentido oposto aos aceleradores, aumentando o tempo de hidratação da pasta de cimento de modo a permitir a colocação da pasta na profundidade desejada em tempo seguro. Esses aditivos não interferem na resistência final da pasta, mas apenas na taxa de desenvolvimento desta.

2.2.1.2.3 Dispersante

As pastas de cimento possuem uma alta concentração de partículas sólidas em suspensão na água. As interações entre essas partículas são função da distribuição de cargas na superfície dos grãos de cimento. Os dispersantes ajustam as cargas para se obter as propriedades reológicas desejadas. Otimizar as propriedades reológicas significa possibilitar o bombeio da pasta de cimento a vazões mais altas e com menores

perdas de carga, além de permitir a sua mistura e as propriedades desejadas nas condições de superfície.

2.2.1.2.4 Estendedor

São aditivos que absorvem água ou possuem uma densidade absoluta inferior à do cimento. Os estendedores são utilizados para reduzir o peso específico ou o custo da pasta. O tipo de estendedor a ser misturado será selecionado de acordo com o objetivo de sua utilização e as demais propriedades desejadas.

2.2.1.2.5 Adensante

Os adensantes atuam de maneira contrária aos estendedores, aumentando a o peso específico da pasta. Para isso, são materiais de massa específica superior à do cimento.

2.2.1.2.6 Viscosificante

Necessário para o controle das propriedades reológicas e manutenção dos sólidos em suspensão.

2.2.1.2.7 Controlador de Filtrado

Os controladores de filtrado são utilizados para controlar a quantidade de água perdida pela pasta de cimento por filtração para as formações permeáveis que compõem a parede do poço. Ou seja, seu objetivo é manter a razão sólido/líquido da pasta de cimento durante o tempo de colocação e o tempo de cura. Dessa maneira, evita-se a desidratação prematura da pasta de cimento, mudanças em sua propriedade e o dano induzido em zonas de produção pelo filtrado de cimento.

2.2.1.2.8 Controlador de Gás

A migração de gás consiste na invasão do material gasoso proveniente da formação pelo anular, evento este indesejado. Para evitar o fenômeno, são adicionadas às pastas de cimento aditivos capazes de controlar a migração de gás, sendo os sistemas mais comuns baseados no uso de látex.

2.2.1.2.9 Agentes Expansivos

As pastas de cimento com agentes expansivos têm como característica a compensação da contração volumétrica que ocorre durante a pega do cimento pela expansão de seu material. O tempo de expansão irá depender do aditivo utilizado e da formulação da pasta, podendo levar semanas para expansão completa.

2.2.1.2.10 Sílica

A sílica é utilizada para prevenir a retrogressão da pasta de cimento. Esse fenômeno ocorre devido à deterioração da matriz de cimento pela mudança na estrutura do cimento hidratado a altas temperaturas, superiores a 230°F, ocasionando a perda de resistência à compressão e aumento da permeabilidade da pasta.

2.2.1.3 Colchões lavadores e espaçadores

A completa remoção do fluido de perfuração é de extrema importância para uma melhor aderência do cimento à formação e ao revestimento e, portanto, para o sucesso da operação de cimentação. Para auxiliar nessa remoção, são utilizados colchões lavadores e espaçadores, bombeados antes da pasta de cimento.

Os colchões lavadores são geralmente fluidos poucos viscosos e de baixa densidade, compatíveis com o fluido de perfuração e com a pasta de cimento, que atuam por meio de ação química e mecânica na diluição e remoção do reboco (filme de baixa permeabilidade formado pelos fluidos de perfuração nas paredes do poço).

Os colchões espaçadores são fluidos viscosos e de maior densidade que atuam mecanicamente na remoção do reboco. Esses fluidos são também bombeados à frente da pasta de cimento a fim de evitar sua contaminação com o fluido de perfuração, possibilitando assim que a pasta chegue íntegra no anular do poço. A contaminação da pasta de cimento pelo fluido de perfuração pode causar a gelificação na interface cimento/fluido, reduzindo a eficiência de deslocamento e propiciando a formação de canais de fluido no interior do cimento colocado no anular, prejudicando o isolamento hidráulico deste.

2.2.1.4 Ensaios Laboratoriais

Com base nas informações preliminares das condições de poço às quais a pasta será submetida durante a operação de cimentação, é escolhido o sistema de pasta de cimento a ser utilizado. Essa seleção consiste na definição da densidade e tipo da pasta e dos aditivos a serem misturados e de suas concentrações.

Os aditivos químicos, apesar de terem uma função primária, podem apresentar também funções secundárias, além de interagirem com os demais aditivos incluídos, podendo ter suas propriedades alteradas. Por essa razão e devido à variação da composição da pasta de cimento em função da batelada, testes laboratoriais são imprescindíveis para prever o desempenho da pasta de cimento a ser utilizada. O “Committee on Standardization of Well Cements (Committee 10)” padronizou os ensaios laboratoriais a serem realizados, publicados na “API Spec 10” [5]. Com base nesse documento, foi criado o PROCELAB, procedimento de ensaios utilizados pela Petrobras. [27]

O fluxograma apresentado na Figura 10 contém os principais equipamentos de um laboratório de cimentação e a sequência a ser seguida quando da execução dos testes para a determinação das propriedades físicas da pasta de cimento. [27]

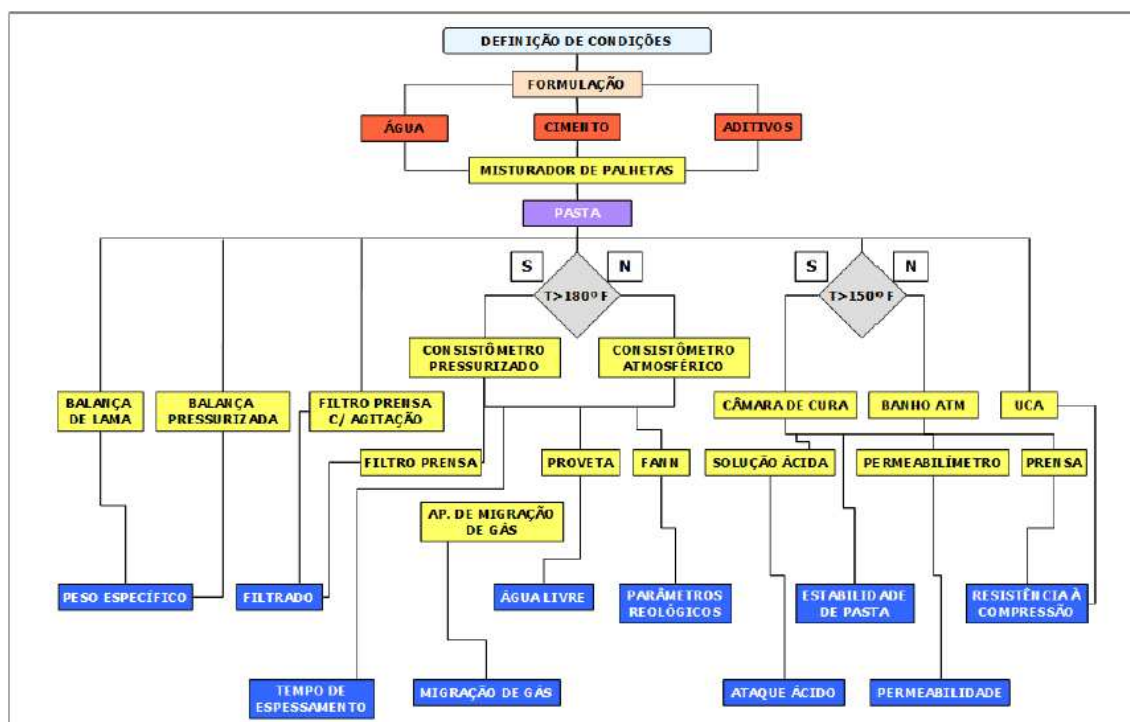


Figura 10 - Fluxograma dos principais testes de pasta [27]

2.2.1.4.1 Peso Específico

O termo peso específico, utilizado no fluxograma, é difundido no meio operacional, no entanto, o termo correto seria massa específica. O ensaio tem como objetivo, portanto, determinar a massa específica aparente (podendo conter uma quantidade de ar aprisionada significativa) e a massa específica real (quantidade de ar aprisionada desprezível) de pastas de cimento.

A massa específica da pasta de cimento deve ser adequada às curvas de geopressões do poço. Quanto menor a massa específica da pasta de cimento, menor será a pressão hidrostática durante a cimentação. Essa característica permite cimentações de extensões mais longas sem a necessidade de ser realizada em mais de um estágio e previne a indução de perda de circulação devido a fratura de formações fracas. Já no caso de poços HP (*High Pressure* – de altas pressões), as pastas de cimento devem ter massa específica mais elevadas para não permitirem o influxo dos fluidos da formação.

2.2.1.4.2 Filtrado

O ensaio determina a perda da fase líquida da pasta de cimento, em condição estática, através de um elemento filtrante. A desidratação da pasta de cimento pode levar a um aumento na reologia da pasta, prejudicando o fluxo no anular, à diminuição do tempo de bombeabilidade da pasta e a um dano na formação.

2.2.1.4.3 Tempo de Espessamento

O tempo de espessamento é definido como o tempo requerido para atingir 100 UC (unidades de consistometria). Esse valor representa o tempo estimado que a pasta de cimento permanece fluida sob determinadas condições de pressão e temperatura. O tempo para atingir 50UC é chamado por sua vez de tempo de bombeabilidade e representa então o tempo disponível para bombeio da pasta de cimento até sua completa colocação no espaço anular.

2.2.1.4.4 Água Livre

O ensaio de água livre prevê a susceptibilidade de uma pasta de cimento segregar, possibilitando assim estimar seu volume total efetivo reduzido. Em poços desviados, a formação de água livre é ainda mais crítica, possibilitando a formação de canais de fluido, uma vez que a água será formada na parte superior do poço.

2.2.1.4.5 Parâmetros Reológicos

A reologia estuda as deformações e escoamentos do material. O estudo das propriedades reológicas das pastas de cimento visa avaliar se a pasta de cimento é misturável na superfície e otimizar a eficiência com que a pasta de cimento desloca o fluido no espaço anular sob determinado regime de fluxo e a real pressão exercida sobre as paredes do poço [10]. Nas condições de fundo, é importante que a pasta seja fluida, de forma a minimizar a perda de carga (perdas por fricção) durante o deslocamento da pasta. Por outro lado, deve ter uma reologia que confira estabilidade, impedindo a decantação de sólidos e formação de água livre, e que seja superior ao do fluido que será deslocado.

As principais leituras da pasta analisadas através das propriedades reológicas são:

- Limite de Escoamento (LE): tensão mínima a ser aplicada para que a pasta de cimento entre em movimento

- Viscosidade Plástica (VP): constante de proporcionalidade entre a tensão de cisalhamento (força por unidade de área cisalhante necessária para manter o escoamento dos fluidos) e a taxa de cisalhamento (deslocamento relativo das partículas de fluido relacionado com a distância entre elas) para tensões superiores ao limite de escoamento. De uma maneira simples, deve ser baixa para facilitar sua penetração nos anulares, oferecer boas condições de bombeabilidade e aderir fortemente ao revestimento e à formação. [21]

- Gel Inicial (GI): resistência que a pasta de cimento apresenta para reiniciar o movimento após uma parada de 10 segundos. Indica o grau de tixotropia da pasta de cimento.

- Gel Final (GF): resistência que a pasta de cimento apresenta para reiniciar o movimento após uma parada de 10 minutos. Indica o grau de tixotropia da pasta de cimento.

2.2.1.4.6 Resistência à compressão

O intuito dos ensaios de resistência à compressão é o de determinar a resistência da pasta de cimento endurecida e, portanto, as cargas às quais a pasta de cimento suporta. Pelo método ultrassônico (UCA), é gerado um gráfico contínuo de desenvolvimento de resistência compressiva em função do tempo.

2.2.2 Operação de Cimentação

Normalmente, nas fases iniciais do poço, a cimentação possui função estrutural de estabilizar a cabeça do poço. Já nas fases intermediárias e de produção, o objetivo principal das cimentações é o isolamento hidráulico das formações, garantindo a segurança operacional, sendo, portanto, o foco de análise desse trabalho.

2.2.3 Fatores que Afetam a Qualidade da Cimentação

2.2.3.1 Eficiência de Deslocamento

Para se obter sucesso na operação de cimentação, é necessário que toda a circunferência do anular seja preenchida pela pasta de cimento e que esta desenvolva aderência ao revestimento e à formação (ou ao revestimento anterior, em casos de anulares entre revestimentos). Para isso, é necessária uma eficiente remoção do fluido de perfuração.

O deslocamento incompleto do fluido de perfuração pode levar a falhas na cimentação como a canalização da pasta de cimento, permitindo fluxo indesejado de fluidos da formação pelo espaço anular. A contaminação da pasta de cimento, sofrendo alterações nas propriedades desejadas, pode ser também uma consequência da falha de remoção do fluido de perfuração.

É observado que as melhores eficiências de remoção do fluido de perfuração são alcançadas quando a pasta de cimento é deslocada a altas vazões, independentemente do tipo de regime de fluxo [26]. Além disso, outros fatores podem influenciar nessa eficiência como a manutenção da hierarquia reológica dos fluidos, volumes de colchões bombeados, dentre outros.

2.2.3.2 Centralização do Revestimento

Centralizadores são acessórios utilizados na coluna de revestimento a fim de se obter uma área de anular uniforme, em que a distância do revestimento a todos os lados da parede do poço seja similar. Dessa forma, evita-se a formação de canais preferenciais para a passagem de fluido na parte mais larga do anular, ou seja, com menor resistência ao fluxo.

Em poços desviados, a utilização de centralizadores na coluna de revestimento se torna ainda mais crítica, devido a uma tendência de descentralização do revestimento pelo efeito gravitacional.

2.2.3.3 Volume da Pasta de Cimento

O volume de pasta de cimento a ser bombeado deve ser calculado de acordo com a altura de cimento que se deseja obter no anular. Para esse cálculo, é necessário

também conhecer o diâmetro do poço, que está relacionado ao diâmetro da broca utilizada durante a perfuração da fase. No entanto, dependendo do fluido de perfuração, formações perfuradas e outros parâmetros da operação de perfuração, o poço pode sofrer arrombamentos, ficando com um diâmetro real maior do que o previsto. Em alguns casos, utiliza-se a informação do perfil caliper para a informação do diâmetro real do poço, mas nem sempre se dispõe dessa informação.

Um volume de pasta de cimento menor do que o necessário poderá fazer com que formações permeáveis, que deveriam ser isoladas hidráulicamente, fiquem expostas devido ao topo do cimento mais profundo do que o desejado. A fim de evitar esse problema, o volume de pasta de cimento a ser bombeado é calculado considerando o volume nominal do anular acrescido de um determinado excesso. O excesso de pasta bombeado pode ainda compensar situações de perda de circulação da pasta de cimento para formação.

2.2.3.4 Propriedades do fluido de perfuração

Para facilitar a remoção do fluido de perfuração, é necessário que este tenha certas características que, muitas vezes, são diferentes das características necessárias para a perfuração do poço. É indicado então o condicionamento do fluido dentro do poço antes do início do bombeio da pasta de cimento, de maneira a obter as seguintes propriedades: [26]

- Baixos valores de viscosidade plástica e limite de escoamento, assim como baixos valores da força gel (baixa característica tixotrópica);
- Baixo peso específico de modo a facilitar o deslocamento da lama pela pasta de cimento;
- Baixo filtrado para prevenir a formação de um reboco espesso e de difícil remoção;
- Composição química compatível com a da pasta de cimento.

2.2.3.5 Uso do plugue de topo e de fundo

Os plugues são utilizados para separar os fluidos durante o deslocamento da pasta de cimento. O plugue de fundo é utilizado à frente da pasta ou do colchão, preferencialmente após o colchão e antes da pasta, afim de utilizar o colchão como

material de sacrifício e privilegiando a preservação da pasta. Sua utilização evita contaminações com o fluido de perfuração. O plugue de topo, deslocado imediatamente após a pasta de cimento, previne contaminações provocadas pelo fluido de deslocamento. [26]

2.2.3.6 Perdas de circulação

A perda de circulação consiste na invasão do fluido para a formação através de fraturas naturais ou induzidas ou em formações depletadas. As perdas monitoradas para o controle de qualidade da cimentação podem ocorrer durante a descida do revestimento ou da circulação para cimentar, situação em que é perdido fluido de perfuração, ou durante a operação de cimentação. Para as perdas ocorridas durante a operação de cimentação, é difícil identificar com precisão qual o fluido presente no poço que invadiu a formação, podendo ser fluido de perfuração, colchões ou a pasta de cimento.

A perda dos colchões poderá prejudicar o efeito esperado para o bombeio destes fluidos, dificultando a remoção do fluido de perfuração e aumentando a possibilidade de contaminação da pasta de cimento.

A perda de pasta de cimento para a formação irá acarretar em topo do cimento abaixo do planejado, podendo prejudicar a obtenção dos isolamentos requeridos. A taxa de perda de fluido nas etapas anteriores à cimentação pode ajudar a prever o volume de excesso de pasta de cimento a ser bombeado a fim de compensar o volume perdido.

2.3 Avaliação de qualidade da cimentação através de perfis acústicos

As ferramentas de perfilagem acústicas são comumente utilizadas para avaliar a qualidade da cimentação de poços de petróleo. Essas análises não permitem uma medida direta da capacidade de isolamento do anular pelo cimento, mas fornecem um indicativo da sua presença e de aderência cimento-revestimento e cimento-formação. A partir dessas informações é possível inferir a obtenção do isolamento hidráulico.

As ferramentas ultrassônicas existentes no mercado se dividem em dois grupos, de acordo com a frequência de emissão do sinal acústico:

- Ferramentas sônicas,
- Ferramentas ultrassônicas.

As leituras das ferramentas disponíveis no mercado são complementares e todas elas têm suas vantagens e limitações. Dessa forma, a combinação de ferramentas em uma mesma corrida de perfilagem permite uma maior confiabilidade e melhor entendimento da qualidade da cimentação no anular.

2.3.1 Ferramentas Sônicas

2.3.1.1 Cement Bond Log (CBL)

A ferramenta CBL é composta por um emissor e um receptor a uma distância de 3 pés. O emissor, acionado eletricamente, vibra gerando ondas sonoras que irão ser emitidas omnidirecionalmente, isto é, para todas as direções. O sinal recebido é então uma composição das ondas que caminharam por todas as direções e em todas as interfaces do poço, conforme Figura 11 e Figura 12.

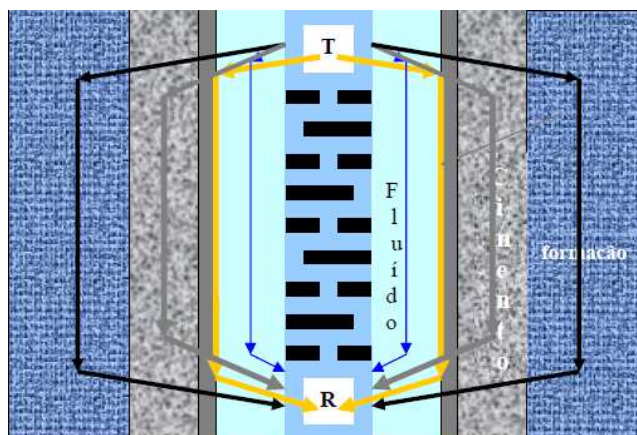


Figura 11 - Trajetória de energia acústica entre transmissor e receptor a 3 pés [11]

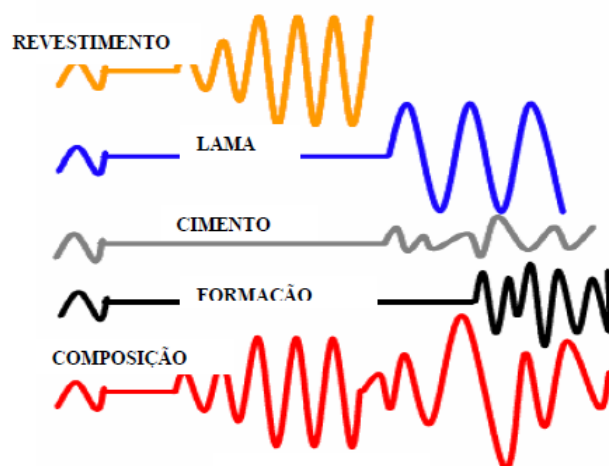


Figura 12 - Onda composta recebida no transdutor a 3 pés [11]

É esperado que o primeiro sinal recebido seja referente à onda que caminhou pelo revestimento, devido à maior velocidade de propagação do sinal acústico no aço que nos demais materiais normalmente encontrados. A leitura do CBL será então a amplitude do primeiro pico recebido e fornecerá um indicativo da aderência cimento-revestimento. Amplitude alta indica que pouca energia foi dissipada e entende-se então que o material no anular é fluido enquanto que, por outro lado, amplitude baixa indica grande dissipação de energia, inferindo-se a existência de cimento de boa qualidade no anular.

Em alguns casos, como em formações salinas e outras muito consolidadas, a velocidade de propagação pode ser maior do que no revestimento, gerando uma interferência no sinal recebido e podendo dificultar a interpretação. O mesmo pode ocorrer em situações de revestimentos concêntricos.

Existem ainda outros tipos de interferência que podem ocorrer, que variam de acordo com as características do poço, resultando em leituras anômalas de amplitude do CBL. No entanto, não é objetivo deste trabalho aprofundar em detalhes de interpretação dos perfis de avaliação da cimentação.

Existem algumas ferramentas no mercado, como o MSIP e o ASLT, providas pelas Schlumberger, e o SBT, provida pela Baker Hughes, que são variações do CBL convencional. Para as ferramentas da Schlumberger, há algumas pequenas diferenças na forma de aquisição das informações. O SBT por sua vez, tem diferenças maiores, já que os transdutores são acoplados a patins que correm encostados no revestimento e a frequência de emissão do sinal está na faixa ultrassônica. No entanto, em geral, as leituras finais são similares às descritas e a forma de interpretação dos dados é similar.

2.3.1.2 Variable Density Log (VDL)

O VDL é uma leitura proveniente do mesmo emissor utilizado no CBL, no entanto agora o receptor está localizado a 5 pés do emissor. Assim como para o CBL, a leitura no receptor será de uma composição do sinal acústico que caminhou em todas as direções e em todas as interfaces do poço, conforme Figura 13

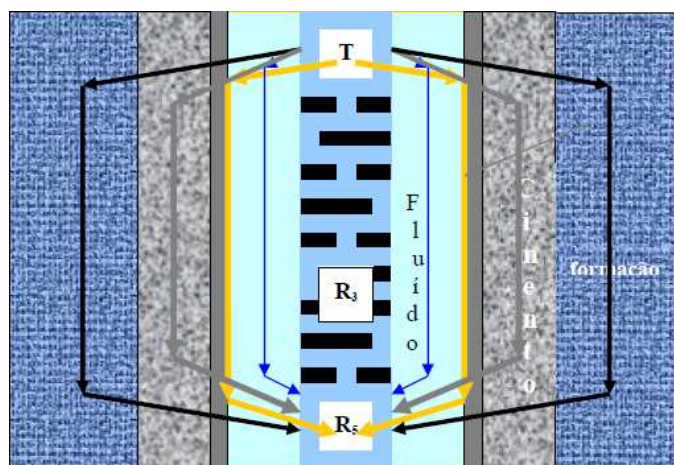


Figura 13 - Trajetória de energia acústica entre transmissor e receptor a 5 pés [11]

Para o VDL será utilizado todo o trem de ondas recebido e, portanto, além da análise cimento-revestimento, é possível analisar a aderência cimento-formação. O VDL é apresentado na forma de microsismograma, em que a onda recebida é transformada de maneira que os vales são pintados de branco e os picos em uma escala de cinzas a depender da amplitude do mesmo, até atingir a cor preta, para leituras altas de amplitude, conforme Figura 14.

Normalmente, os primeiros picos recebidos são referentes às ondas que caminharam pelo revestimento. A característica dos sinais recebidos após o sinal do revestimento irá depender do material presente no anular. No caso de anular livre, sem presença de cimento, o sinal será proveniente do fluido no interior do revestimento, e tem como características sinais fortes e retilíneos. Já em casos de revestimentos bem cimentado, o sinal será proveniente das ondas que percorreram a formação e, por último, as ondas que viajaram pelo fluido no interior do poço.

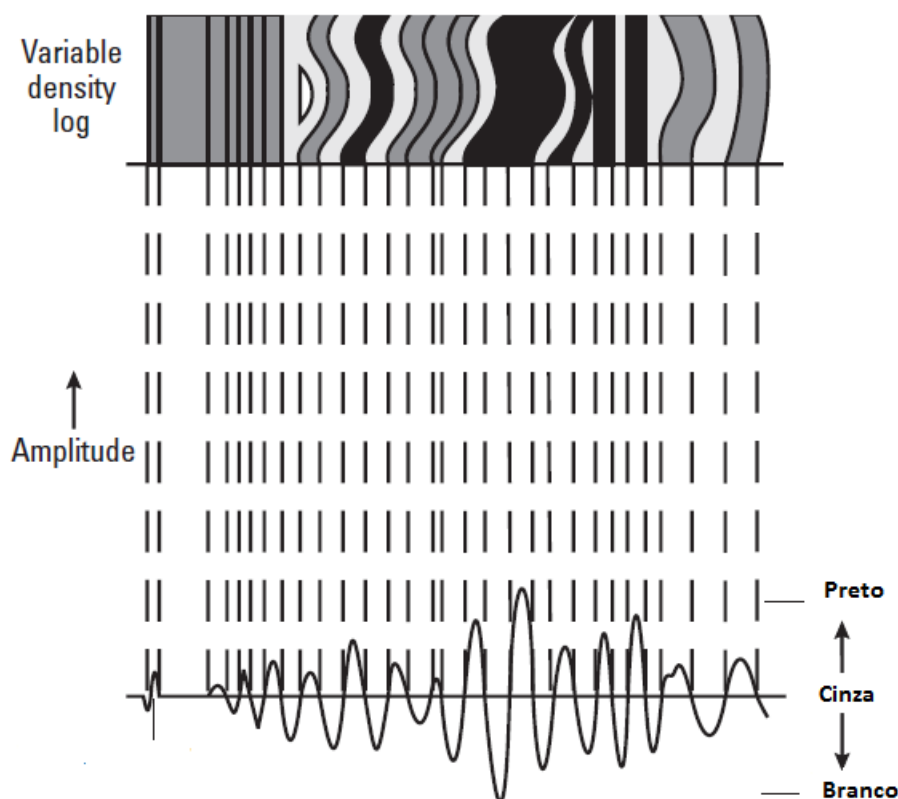


Figura 14 – Microsismograma VDL (Adaptado de [4])

2.3.2 Ferramentas Ultrassônicas

As ferramentas ultrassônicas são compostas por um transdutor rotativo a partir do qual é inferida a impedância acústica do material no anular. O transdutor emite um sinal ultrassônico com frequência suficiente para ressonar o sinal na parede do revestimento, conforme Figura 15. A proporção do sinal que irá refletir ou refratar em cada uma das interfaces atingidas, é determinada pelo coeficiente de reflexão que, conforme Equação 1, depende da impedância dos materiais presente nos dois lados da interface. A Tabela 1 apresenta valores de impedância característicos de alguns materiais.

As ferramentas ultrassônicas, ao contrário das ferramentas sônicas, fornecem um mapa de cimento de alta resolução, já que são obtidos de 60 a 100 pontos em uma circunferência do poço. Essa propriedade permite a identificação de canais no anular cimentado.

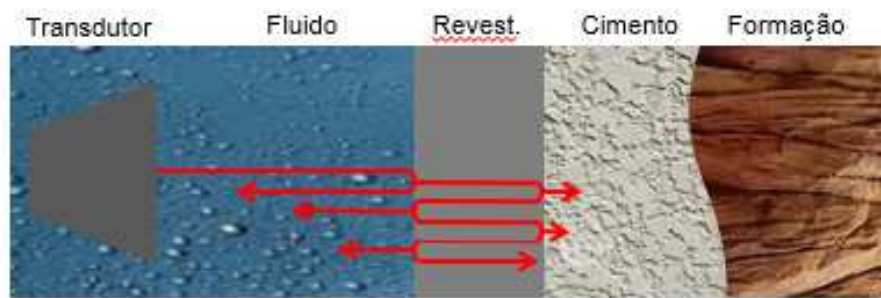


Figura 15 - Princípio de funcionamento das ferramentas ultrassônicas

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \quad (1)$$

Tabela 1 - Impedância acústica características dos materiais (Traduzido de [13])

Material	Impedância Acústica (Mrayl)
Água	1.50
Gás	0.10
Aço (Revestimento)	46.00
Fluido de perfuração 12 lb/gal	2.16
Fluido de perfuração 15 lb/gal	2.70
Fluido de perfuração 17 lb/gal	3.06
Espuma classe C 9 lb/gal (250 psi)	2.19
Espuma classe C 9 lb/gal (1000 psi)	2.69
Cimento 13 lb/gal (500 psi)	3.37
Cimento 13 lb/gal (2000 psi)	4.42
Cimento 16.5 lb/gal (500 psi)	4.38
Cimento 16.5 lb/gal (2000 psi)	5.62

2.3.3 Perfis de Cimentação

A partir dos dados coletados em uma operação de perfilagem, é gerado um perfil que retrata a condição da cimentação, metro a metro, por toda a extensão perfilada, geralmente contendo informações de toda a extensão cimentada. A Figura 16 apresenta um exemplo do retrato da cimentação obtido através da perfilagem.

As curvas presentes no perfil dependem das ferramentas corridas para perfilagem. O resultado final obtido pela interpretação dos dados, no entanto, é sempre o mesmo: a inferência de isolamento hidráulico em cada trecho, para toda a extensão perfilada, obtida pelas características observadas nas leituras das ferramentas que indicam a qualidade da cimentação.

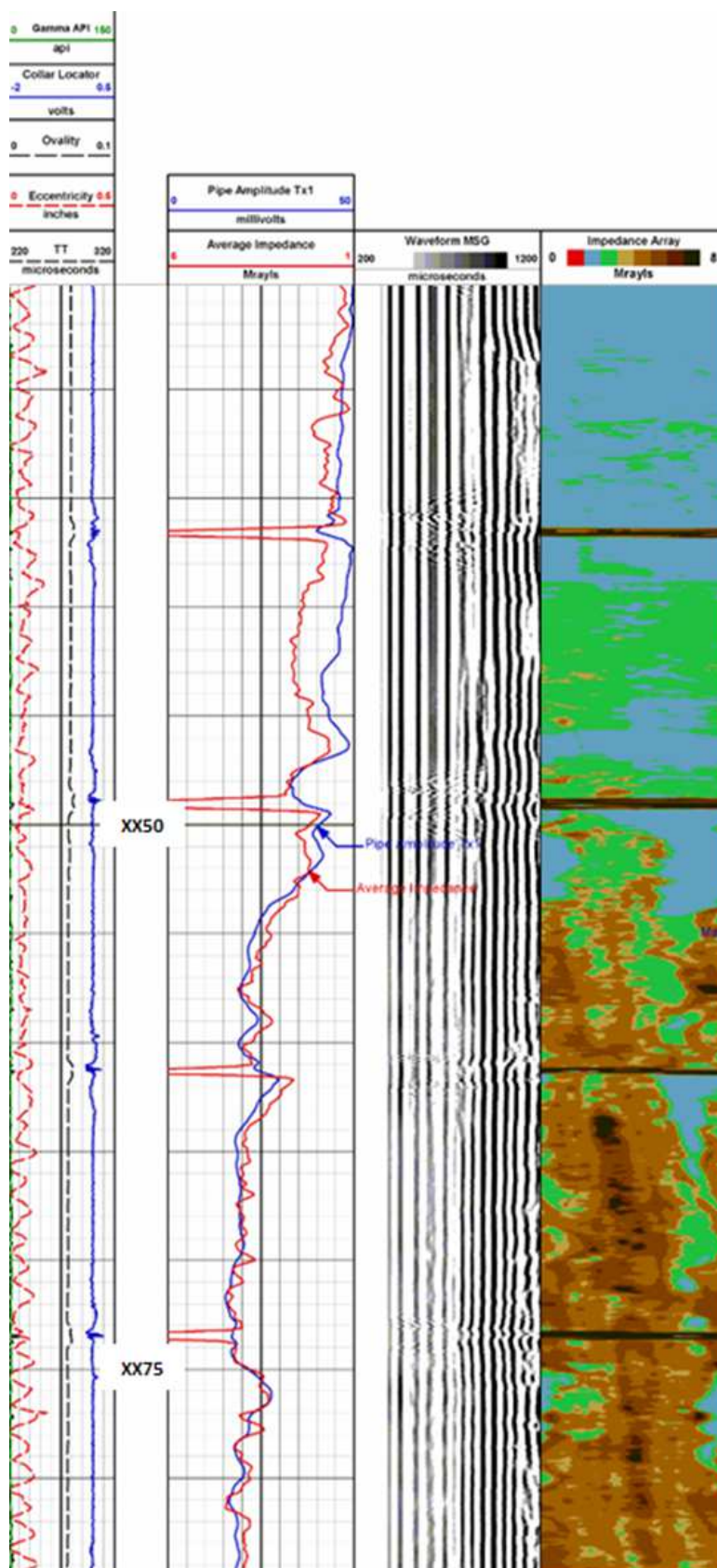


Figura 16 - Perfil de Avaliação da Qualidade da Cimentação [1]

3 O Conjunto de Dados

Um banco de dados começou a ser construído em 2013 com os resultados da qualidade da cimentação dos poços de petróleo offshore da Petrobras em que foram corridos perfis acústicos para avaliação da qualidade da cimentação. Cada um dos registros é associado à qualidade da cimentação de uma coluna de revestimento de uma das fases do poço.

As cimentações avaliadas são classificadas segundo os critérios apresentados na Tabela 2. A base para a classificação da cimentação é a obtenção de isolamento hidráulico requerido, ou seja, boa qualidade da cimentação nas profundidades de interesse, a fim de atender aos critérios para construção das barreiras de segurança e para produção do poço. Para as cimentações em que qualquer isolamento hidráulico requerido não é obtido, é efetuada uma operação de correção da cimentação, que tem custo elevado e alto grau de complexidade.

Tabela 2 – Classificação da qualidade da cimentação

Classificação	Crítérios
Excelente	Atingiu todos os isolamentos requeridos. Inferência de isolamento hidráulico no mínimo em 70% do trecho cimentado.
Boa	Atingiu todos os isolamentos requeridos. Inferência de isolamento hidráulico de 30% a 69% do trecho cimentado.
Média	Atingiu todos os isolamentos requeridos. Inferência de isolamento hidráulico em menos de 30% do trecho cimentado.
Ruim	Não atingiu um ou nenhum dos isolamentos requeridos. Apesar de pelo menos um isolamento hidráulico não ter sido obtido, o perfil indica presença de cimento e boa qualidade da cimentação em outros intervalos cimentados.

Classificação	Critérios
Péssima	Não atingiu nenhum dos isolamentos requeridos. Características de revestimento livre em todo o trecho perfilado.

Foram retiradas das análises os resultados das perfilagens em que há dúvidas de interpretação devido a qualidade dos dados obtidos. O banco de dados, ao qual iremos referenciar como BD1, conta com 461 registros de perfilagem, divididos nas classes conforme Figura 17.

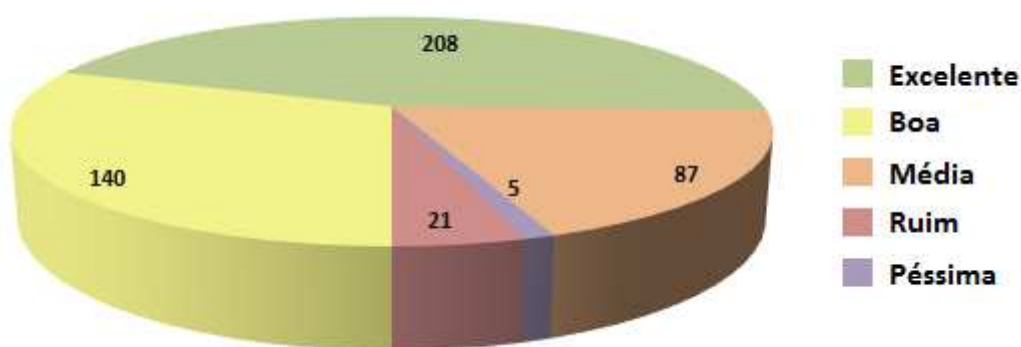


Figura 17 – Classificação da qualidade da cimentação BD1 – 5 classes

A fim de diminuir o desbalanceamento entre as classes, ou seja, classes com poucos registros em comparação com as demais, os registros classificados como “ruim” e como “péssimo” foram unificados, conforme Figura 18. O problema foi estudado também de maneira simplificada, com apenas duas classes: isolamento obtido, composto pelas qualidades “excelente”, “boa” e “média”, e isolamento não obtido, constituído pelas qualidades “ruim” e “péssima”, apresentado na Figura 19.

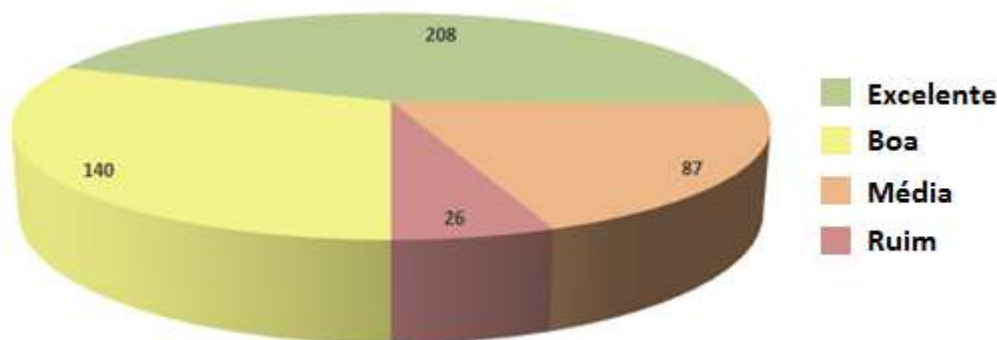


Figura 18 – Classificação da qualidade da cimentação BD1 – 4 classes

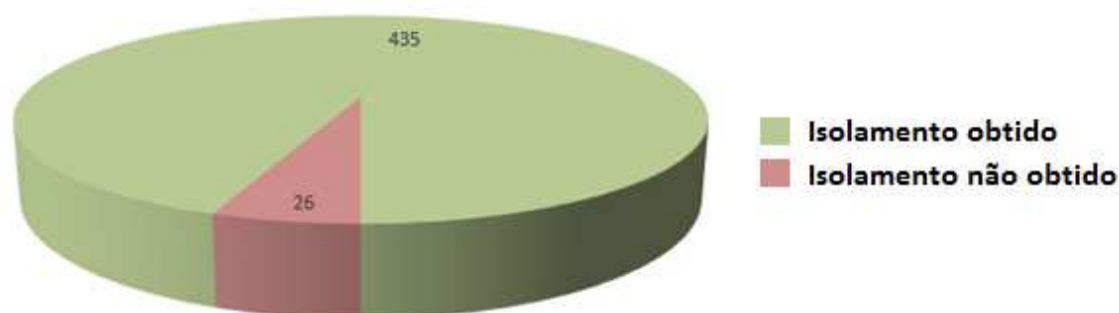


Figura 19 – Classificação da qualidade da cimentação BD1 – 2 classes

Como o banco de dados estudado é acumulativo, ou seja, a cada nova perfilagem são acrescentados os dados referentes àquela cimentação, com o passar do tempo foram identificadas novas variáveis que poderiam enriquecer o banco, não acrescentados seus valores para os registros já realizados. Além disso, durante esse trabalho, foram acrescentadas ainda 16 novas variáveis, relacionadas principalmente aos resultados dos testes de laboratório da pasta de cimento e à sua composição. A lista de todas as 47 variáveis presentes no BD1 é apresentada no APÊNDICE A.

O primeiro passo desse estudo foi a busca na base da Petrobras das informações faltantes, que totalizavam 1667 dados de 17 variáveis quando do início deste trabalho. A Tabela 3 apresenta as variáveis em que, mesmo após a busca e complemento da base de dados, possuem informações ausentes em relação aos 461 registros, totalizando 464

dados ausentes de 23 variáveis. Apenas 196 registros contêm informações de todas as variáveis.

Tabela 3 – Lista de valores ausentes em BD1

Variável	Valores ausentes
Extensão centralizada	22
Diâmetro do centralizador	22
Número de centralizadores	22
Tipo de fluido	3
Fabricante do fluido	3
Densidade do fluido	1
Viscosidade plástica do fluido	4
Limite de escoamento do fluido	5
Leitura 0300 do fluido	14
Leitura 03 do fluido	18
Temperatura de circulação do poço	3
Filtrado	124
Água livre	33
Consistometria (100 UC)	5
Resistência compressiva 24h	125

Variável	Valores ausentes
Leitura 0300 da pasta	7
Leitura 03 da pasta	7
Gel inicial	2
Gel final	2
Viscosidade plástica da pasta	10
Limite de escoamento da pasta	10
Presença de aditivos	2
Vazão de deslocamento	20

Os resultados encontrados utilizando as informações presentes em BD1 não foram satisfatórios, conforme apresentado posteriormente nos resultados deste trabalho, na Seção 5.1. Dessa forma, decidiu-se criar um novo banco de dados, o qual nomearemos de BD2, composto por 247 registros e 16 variáveis, listadas no APÊNDICE B. O número de registros de BD2 é inferior ao de BD1 devido à necessidade de busca completa das informações para sua construção. As variáveis de BD2, com exceção da variável que codifica o registro e daquela que contém a classificação deste, são binárias, indicando a ocorrência ou não de determinado aspecto da operação de cimentação. Alguns desses aspectos são retratados em BD1 por variáveis similares, mas que contêm valores numéricos e não apenas a sua ocorrência como em BD2. A principal diferença entre as variáveis dos dois bancos de dados é que, em BD1 retratam as medidas obtidas no planejamento e na operação de cimentação, enquanto que em BD2 as variáveis retratam a ocorrência de desvios observados na operação. Em BD2 não temos dados ausentes. A divisão dos registros foi feita em 4 classes, conforme critérios presentes na Tabela 2, juntando as classes “Ruim” e

“Péssima”, conforme Figura 20. Foi utilizado o mesmo artifício anterior de simplificação em um problema de apenas 2 classes, apresentado na Figura 21.

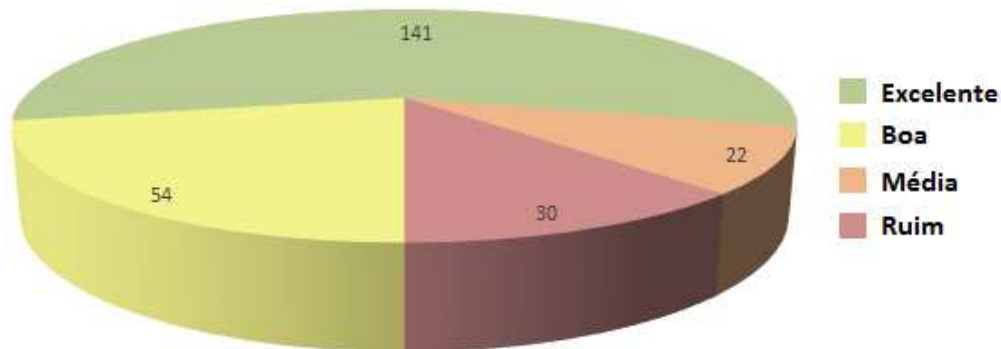


Figura 20 – Classificação da qualidade da cimentação BD2 – 4 classes

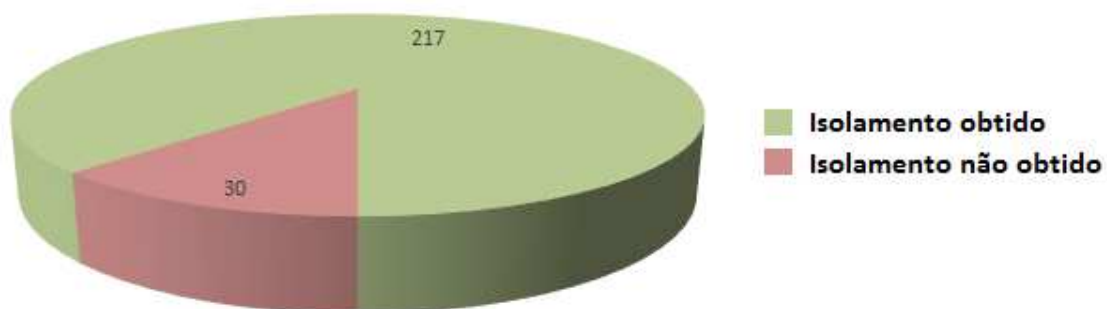


Figura 21 – Classificação da qualidade da cimentação BD2 – 2 classes

Foram analisados ainda os resultados de um terceiro banco de dados, BD3, no qual estão presentes todas as variáveis de BD2 e algumas variáveis de BD1, totalizando 31 atributos, conforme APÊNDICE C. A seleção dos atributos de BD1 para compor BD3 considerou as seguintes premissas:

- Eliminar atributos com muitos dados ausentes;
- Considerar principalmente informações referentes à operação de cimentação, já que os resultados de BD2 se mostraram superiores aos de BD1, conforme apresentado a seguir, na Seção 5. Dessa forma, foram

eliminadas variáveis que descrevem a localização do poço, a fase associada, o fabricante do fluido, dentre outras;

- Eliminar variáveis controladas pela operadora e que, de acordo com o conhecimento histórico, nunca apresentaram problemas para a qualidade final da cimentação, mas são características do cenário ao qual são expostas, como a densidade da pasta de cimento, a temperatura de circulação do poço, dentre outras.

Havia dados conforme BD1 e BD2 de 240 registros, dos quais criou-se BD3, a partir da unificação de suas informações. A classificação dos registros em 4 classes e em 2 classes são apresentadas, respectivamente, na Figura 22 e na Figura 23. Assim como em BD2, em BD3 não estão faltando dados.

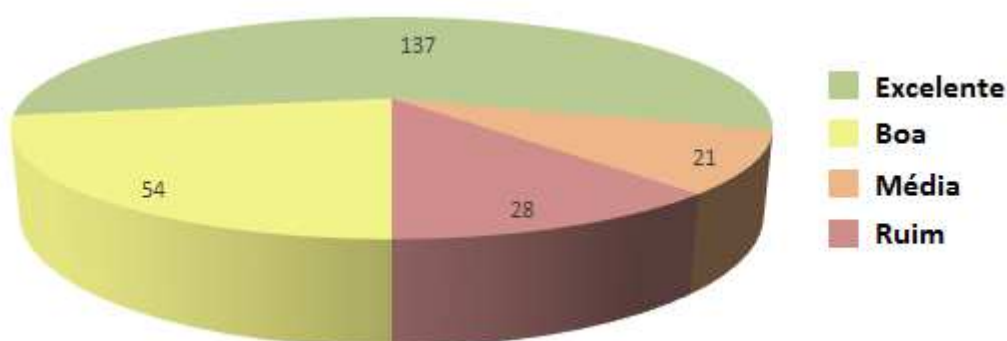


Figura 22 – Classificação da qualidade da cimentação BD3 – 4 classes

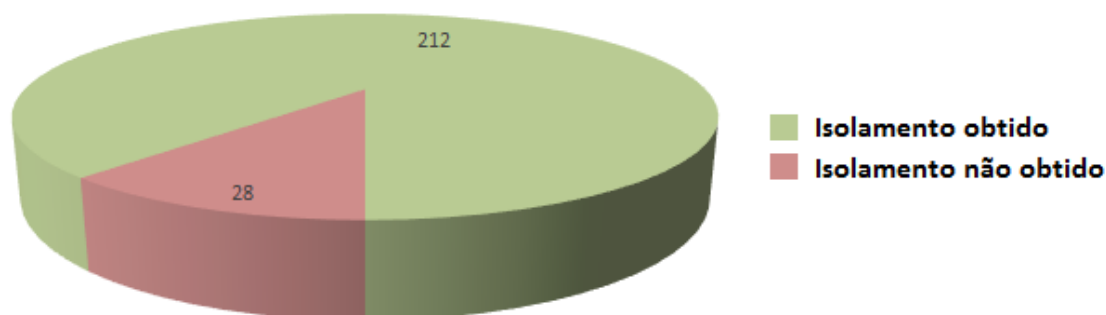


Figura 23 – Classificação da qualidade da cimentação BD3 – 2 classes

O APÊNDICE D apresenta um resumo das características do poço, do planejamento e da operação de cimentação selecionadas em cada um dos três bancos de dados analisados. Algumas características, como é o caso da perda de circulação, estão presentes de forma diferente nos bancos de dados, podendo ser numérico, que indique o valor daquela medida, ou binário, indicando o atendimento do critério.

3.1 Análise dos Dados

A fim de entender as variáveis presentes nos bancos de dados e observar as características gerais de comportamento, foram realizadas as análises estatísticas apresentadas a seguir.

3.1.1 Influência da companhia de cimentação

Dentre as análises estatísticas realizadas, a primeira delas foi a comparação do desempenho das companhias de serviço responsáveis pela cimentação. Conforme apresentado na Figura 24, a análise de falhas por companhia indica que não há diferenças significativas na eficiência das três prestadoras de serviço.

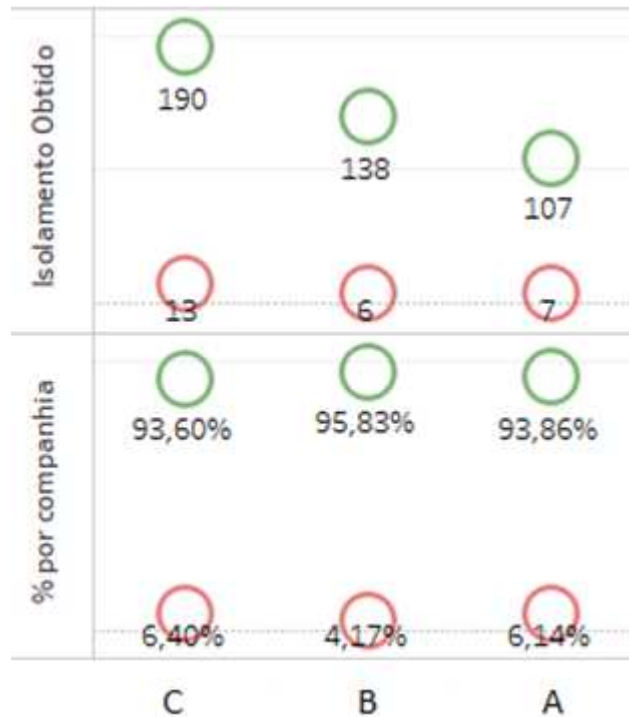


Figura 24 – Análise da qualidade da cimentação por companhia de serviço

3.1.2 Influência da localização do poço

Os poços offshore da Petrobras são divididos de acordo com a região em que estão localizados, em 7 regiões distintas, referenciadas nesse trabalho como P1-P7. A análise de falhas, apresentada na Figura 25, indica que a qualidade da cimentação segue um comportamento similar entre as regiões, a menos da região P6, que se destaca com apenas 1 caso de falha de isolamento hidráulico dentre os 74 analisados, correspondendo a 1,33%. Entende-se que esse resultado melhor do que aos demais se deve ao fato de ser uma região com campos bem conhecidos.



Figura 25 – Análise da qualidade da cimentação por localização do poço

3.1.3 Influência da fase do poço

A fase do poço está relacionada ao diâmetro de poço e de revestimento e, normalmente, as últimas fases, chamadas de “produção” são mais desafiadoras para a cimentação, já que são aquelas com formações de maior permeabilidade e maior possibilidade de perdas de circulação. A análise de falha de isolamento por fase do poço indicou, conforme Figura 26, maior ocorrência nas fases de “produção”, conforme esperado.

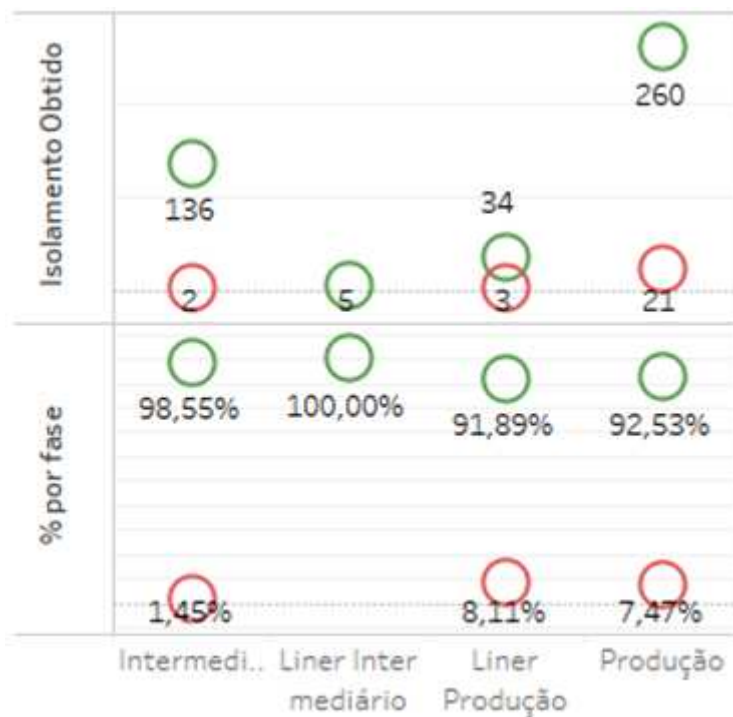


Figura 26 – Análise da qualidade da cimentação por fase do poço

3.1.4 Influência da fase do poço em relação à localização

A fim de buscar a identificação da localização dos poços com falha na cimentação em fases “intermediárias”, realizou-se a análise correlacionando essas duas características do poço, sendo identificado que os dois poços com falha estão localizados em P2, conforme apresentado na Figura 27. Em tentativas de melhorar a resposta de predição dos modelos estudados, foram retirados então dos bancos de dados as fases intermediárias, permanecendo apenas aquelas pertencentes a P2. No entanto, não foram obtidos resultados relevantes.

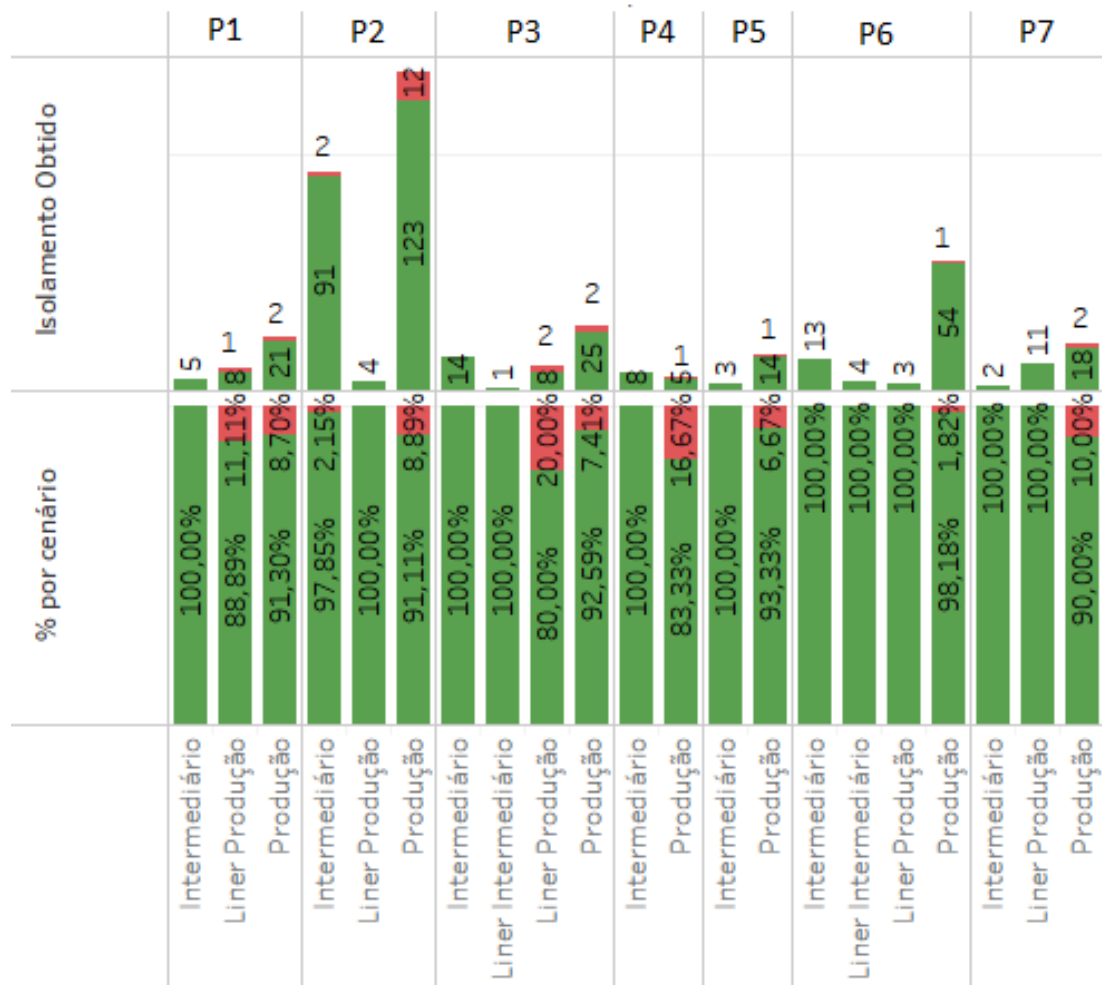


Figura 27 – Análise da qualidade da cimentação por cenário considerando localização e fase do poço

3.1.5 Influência do fluido de perfuração

O fluido presente no poço antes da cimentação pode ser de base aquosa ou base não-aquosa, que será selecionado de acordo com os parâmetros de perfuração desejados. A Figura 28 indica uma diferença significativa de poços cimentados com fluido de base não-aquosa, 384 registros, em relação aqueles com fluido de base aquosa, 48 registros. No entanto, não é observada diferença significativa na qualidade da cimentação considerando essa variável.

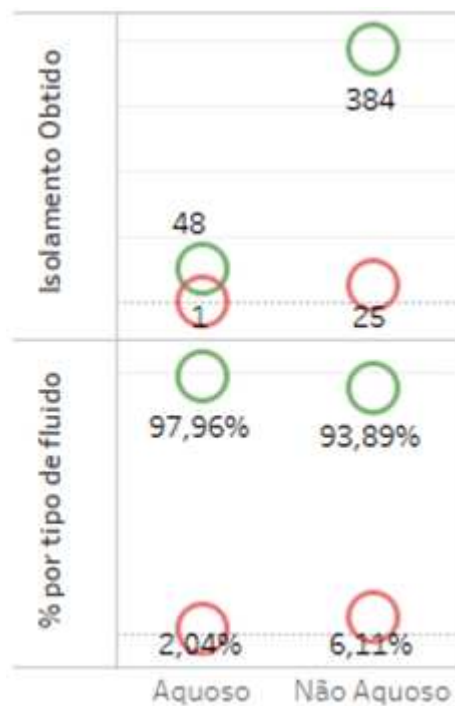


Figura 28 – Análise da qualidade da cimentação por tipo de fluido de perfuração

3.1.6 Influência do fluido de perfuração em relação do fabricante do fluido

A análise de correlação entre o fabricante e tipo de fluido no poço antes da cimentação e a qualidade obtida, apresentada na Figura 29, indica que os maiores problemas ocorrem com fluidos não-aquosos do fabricante F3. É necessário um trabalho de análise aprofundado de compatibilidade desse fluido com as pastas de cimento utilizadas para verificação.

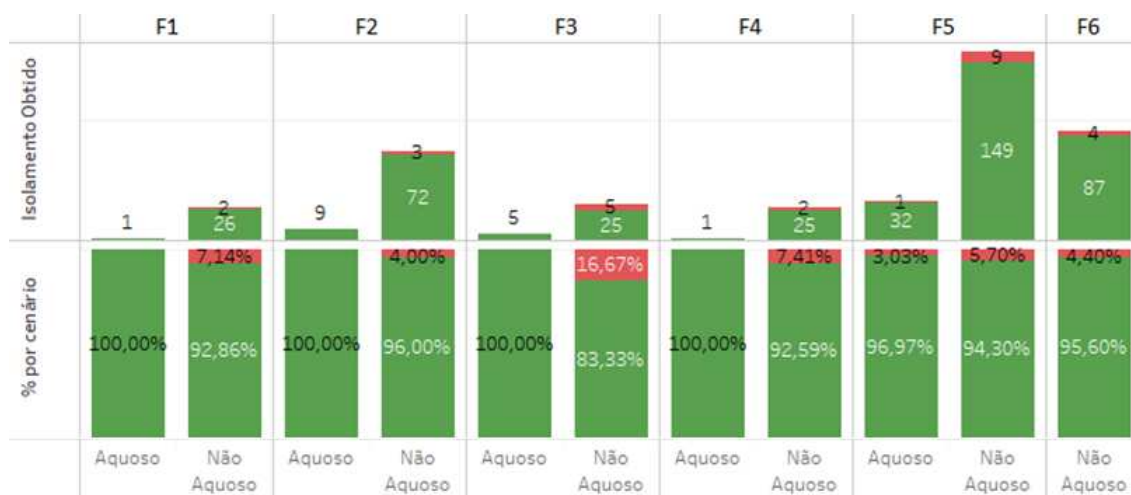


Figura 29 – Análise da qualidade da cimentação por cenário considerando tipo e fabricante do fluido de perfuração

3.1.7 Influência da companhia de cimentação em relação ao fabricante do fluido de perfuração

Para identificar se as falhas na cimentação relacionadas ao fabricante de fluido F3 estão relacionadas a uma companhia de cimentação específica, foi gerado o gráfico apresentado na Figura 30. Foi verificado que não há ocorrências de falha com a companhia de cimentação B, apenas com A e C. Portanto, a análise de compatibilidade entre fluido e pasta de cimento sugerida no item anterior, deve ser priorizada para essas duas companhias de cimentação.

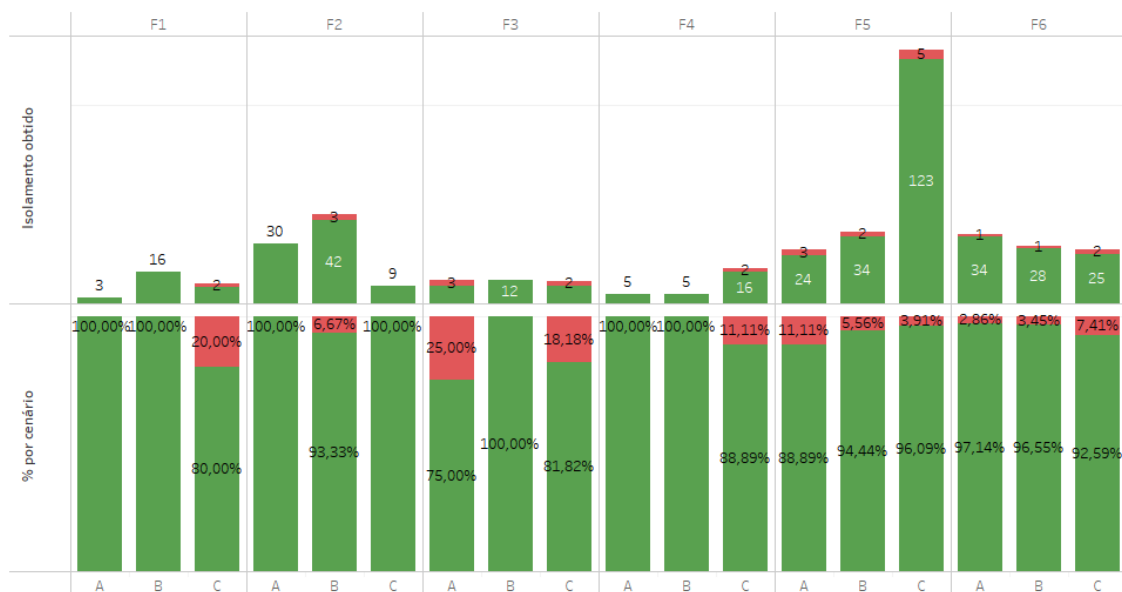


Figura 30 – Análise da qualidade da cimentação por cenário considerando fabricante do fluido de perfuração e companhia de cimentação

3.1.8 Influência da presença de aditivo químico

A análise apresentada na Figura 31, indica que não foi encontrada relação significativa entre falha de cimentação e a presença de determinado aditivo na pasta de cimento.



Figura 31 – Análise da qualidade da cimentação por presença de aditivo químico na pasta de cimento

3.1.9 Influência da inclinação do poço

Intuitivamente, espera-se que quanto maior a inclinação do poço, maiores as dificuldades de cimentação, devido à maior possibilidade de descentralização do revestimento, facilitando a formação de canais na pasta de cimento. No entanto, as análises apresentadas na Figura 32 indicam que não há um histórico de maior ocorrência de falhas na cimentação em poços de maior inclinação.



Figura 32 – Análise da qualidade da cimentação por faixa de inclinação do poço

3.1.10 Influência da companhia de cimentação em relação à inclinação do poço

A análise de falhas de cimentação considerando a inclinação do poço foram estendidas para a relação com companhia de cimentação, conforme Figura 33, e também não indica correlações significativas.

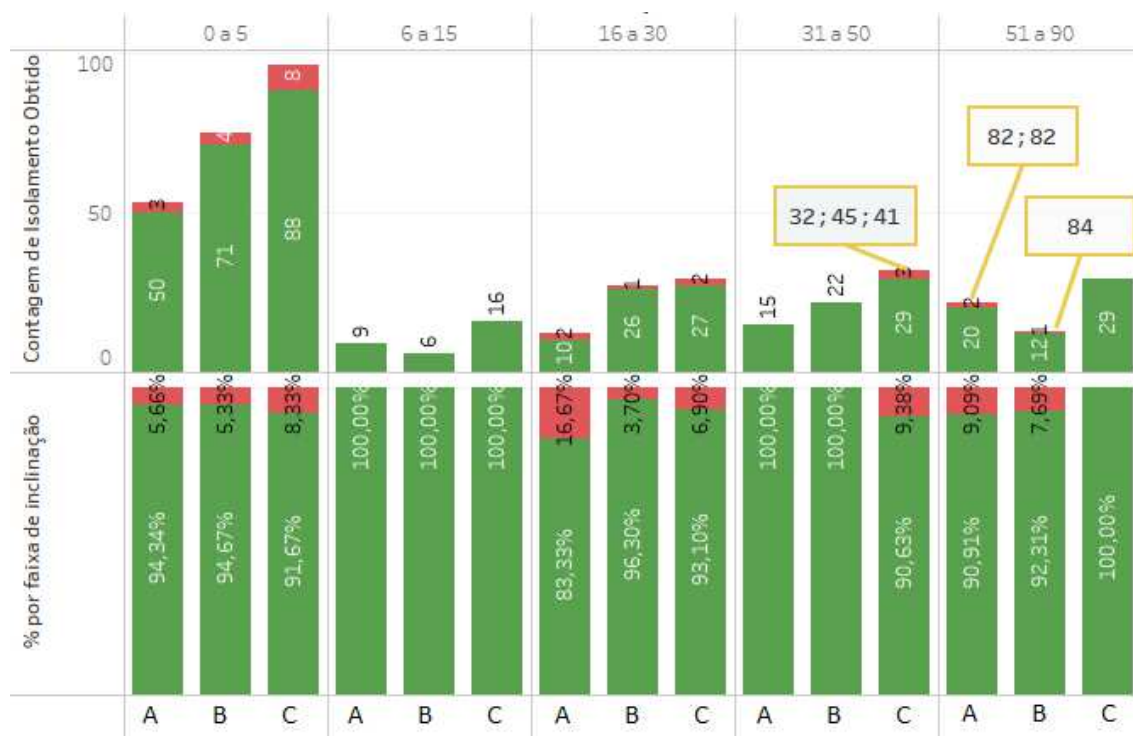


Figura 33 – Análise da qualidade da cimentação por cenário considerando faixa de inclinação do poço e companhia de cimentação

3.1.11 Influência da extensão cimentada

A análise de ocorrência de falhas por extensão cimentada, apresentada na Figura 34, indica, conforme esperado, que cimentações mais extensas são mais susceptíveis a falhas. Isso ocorre devido à maior possibilidade de contaminação da pasta de cimento pelo fluido de perfuração, o que pode causar mudança em suas propriedades.

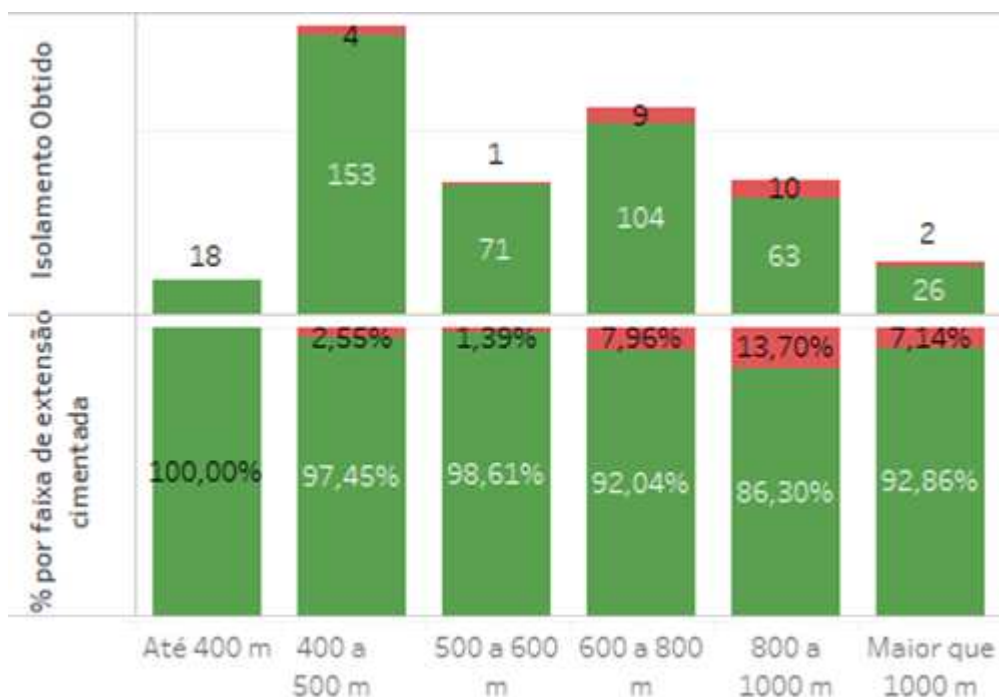


Figura 34 – Análise da qualidade da cimentação por extensão cimentada

3.1.12 Influência da ocorrência de perda de circulação

A análise estatística das variáveis presentes em BD1 indica que o item mais crítico dentre os analisados é a ocorrência de perda durante a operação de cimentação, já que foi o item com maior correspondência com registros de falha na qualidade da cimentação. Na Figura 35 são apresentados gráficos indicando a relação dessa variável com o resultado final (obtenção de isolamento) e com o excesso de pasta de cimento bombeado para o poço. Analisando a primeira coluna do gráfico, temos a relação entre a observação de perda durante a operação de cimentação e a obtenção de isolamento hidráulico. Dentre os poços em que foi observada perda de circulação durante a cimentação, temos 19 ocorrências, correspondendo a 15,2%, de não obtenção de isolamento hidráulico, número relativamente maior que os 2,11% de falha de isolamento entre os poços em que não foi observada perda durante a operação. Dentre os 19 poços com ocorrência de perda e falha de isolamento hidráulico, em 17 deles foi constatado que a perda de cimentação foi maior do que o excesso de pasta de cimento

bombeado para o poço, correspondendo a 16,5% dos poços nesse cenário (perda maior que o excesso e falha de isolamento hidráulico). É esperado que, em casos de volume de perda de pasta de cimento menor do que o excesso de pasta bombeado para o poço, o prejuízo para cimentação não seja tão grande, já que, o volume total de pasta de cimento colocado no anular, desconsiderando o volume perdido, é ainda suficiente para obter o topo de cimento projetado.



Figura 35 - Relação entre perda de circulação, volume de excesso de pasta bombeado e a obtenção de isolamento hidráulico em BDI

3.1.13 Ocorrência de falha na operação de cimentação pela análise dos dados de BD2

Durante fase de pré-processamento de BD2 foram descartadas as informações de 2 dos registros em que foi identificada falha de isolamento hidráulico pela análise do perfil. Para o primeiro deles, a análise de insucesso apresentou como conclusão a utilização de um sistema de pasta (*blend* fornecido por companhia de serviço) que sofreu modificações para acertar suas propriedades. Para o segundo deles, foi constatado como provável causa a contaminação da pasta de cimento em poço de alta inclinação devido à utilização do plugue de fundo antes do colchão, e não entre colchão e a pasta de cimento como é adotado hoje na companhia. As informações desses registros não representam, portanto, o comportamento de sistemas de pasta e operação normalmente adotados, mas indicam que, para cenários com qualquer variação, podendo esta ser de projeto da pasta de cimento, operação ou poço com configuração ou formações não conhecidas, é de extrema importância a perfilagem da cimentação com ferramentas acústicas de forma a comprovar o sucesso do novo sistema.

Avaliando os demais registros conforme BD2, constatou-se que, nas cimentações em que foi identificada falha na cimentação, pelo menos uma das variáveis selecionadas indica falha na operação de cimentação e sendo, portanto, a causa mais provável de falha de obtenção do isolamento hidráulico requerido. Essa análise válida, portanto, para cenários conhecidos, uma lista de parâmetros operacionais que pode ser utilizada para inferir a obtenção dos isolamentos hidráulicos requeridos sem a necessidade de uma perfilagem da cimentação com ferramentas acústicas.

O APÊNDICE E apresenta a relação dos 30 registros em que foi identificada falha de isolamento hidráulico pela análise do perfil acústico e as respectivas falhas observadas durante a operação de cimentação. Os demais 217 registros presentes em BD2 tiveram os isolamentos hidráulicos requeridos inferidos pela análise do perfil acústico de cimentação, dentre os quais apenas 107 registros atenderam a todos os critérios analisados.

Uma análise dos tipos de falhas mais recorrentes é apresentada na Figura 36. Observa-se que em nenhum dos casos em que foi identificada falha de isolamento, foi observado vazamento pelo shoe-track durante o teste negativo do revestimento contra o plugue de topo. Os itens mais recorrentes de falha de cimentação nos poços estudados são os relacionados às perdas de circulação, tanto antes da cimentação quanto durante a operação e deslocamento da pasta de cimento. Esse item já havia sido observado também nas análises de BD1, conforme Figura 35.

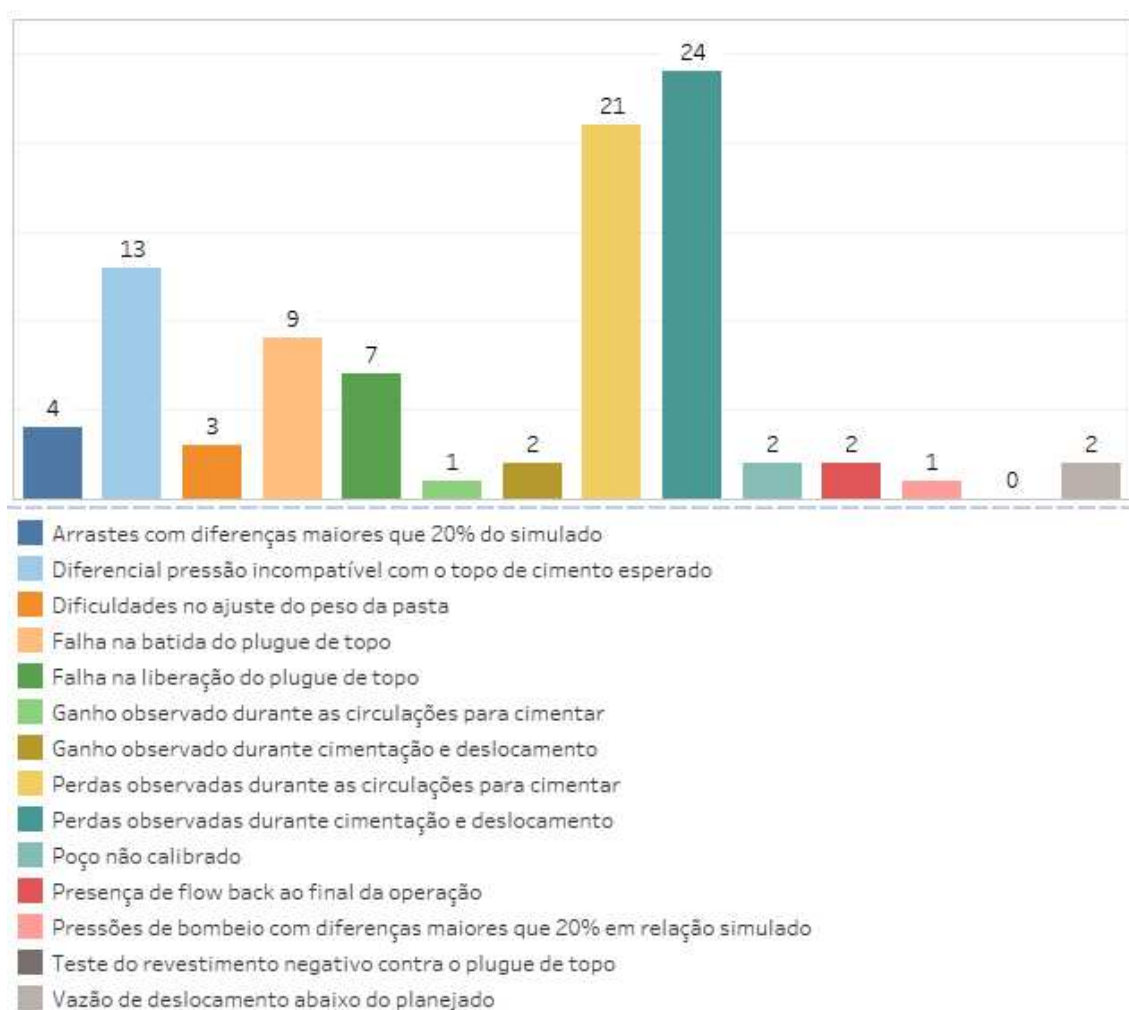


Figura 36 - Recorrência de falhas em BD2 para poços com falha de isolamento hidráulico

4 Modelagem

As análises dos modelos de classificação foram realizadas utilizando a plataforma *Knime (Konstanz Information Miner)*, uma plataforma aberta que permite realizar todo o processo de análise de dados. O *workflow* é exibido graficamente, como um conjunto de nós ligados por setas que indicam a direção do fluxo de dados. Dessa forma, é possível implementar uma tarefa de análise de dados com pouca ou nenhuma programação. A ferramenta oferece mais de 2000 módulos (nós). [20]

Os *workflows* gerados nesse trabalho foram implementados de maneira simples e sem nenhuma programação, sendo, no entanto, suficientes para as análises desejadas.

4.1 Classificação

O objetivo dos algoritmos de classificação é a identificação de padrões a partir dos atributos de forma a identificar a qual classe pré-determinada um determinado registro pertence. O modelo construído será então uma representação do relacionamento entre os atributos e as classes, de forma a possibilitar a previsão de classificação de um novo registro.

4.1.1 Naive Bayes

O classificador Naive Bayes é baseado no teorema Bayesiano, que calcula a probabilidade a posteriori – probabilidade de observar uma determinada classe conhecendo os valores da variável x , $P(C_i|x)$, a partir da probabilidade a priori – probabilidade de ocorrência de uma determinada classe independente de qualquer observação x , $P(C_i)$, e da distribuição de probabilidade condicional – distribuição de probabilidades da variável quando uma determinada classe é observada x , $P(x|C_i)$. [12]

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)} \quad (2)$$

O classificador Naive Bayes considera que a presença (ou ausência) de uma determinada variável não é correlacionada com qualquer outra variável quando a classe

é determinada, ou seja, ele considera que cada variável contribui independentemente para uma determinada classe. [22]

A implementação do modelo no *Knime* é apresentada na Figura 37. O algoritmo implementado utiliza para o cálculo de probabilidade das variáveis nominais o número de registros de acordo com o valor do atributo para cada classe e para o cálculo de probabilidade das variáveis numéricas a distribuição normal (Gaussiana). [20] A configuração do modelo consiste na definição dos seguintes parâmetros, conforme apresentado na Figura 38:

- Coluna que contém a variável alvo, ou seja, as classes (qualidade da cimentação);
- Número máximo de valores únicos para os atributos nominais. Variáveis com número de valores maior do que o definido, são ignoradas durante o treinamento;

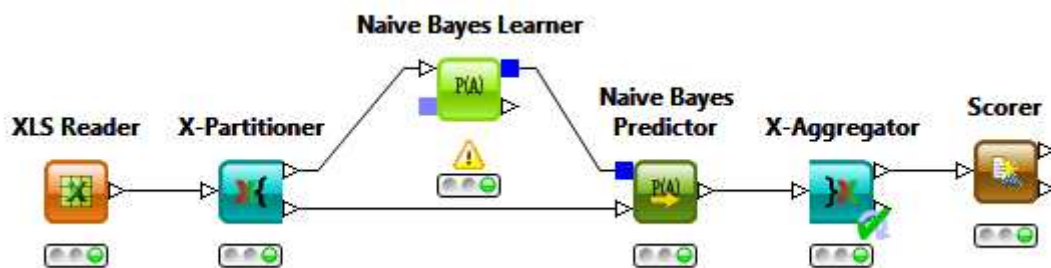


Figura 37 - Workflow Naive-Bayes

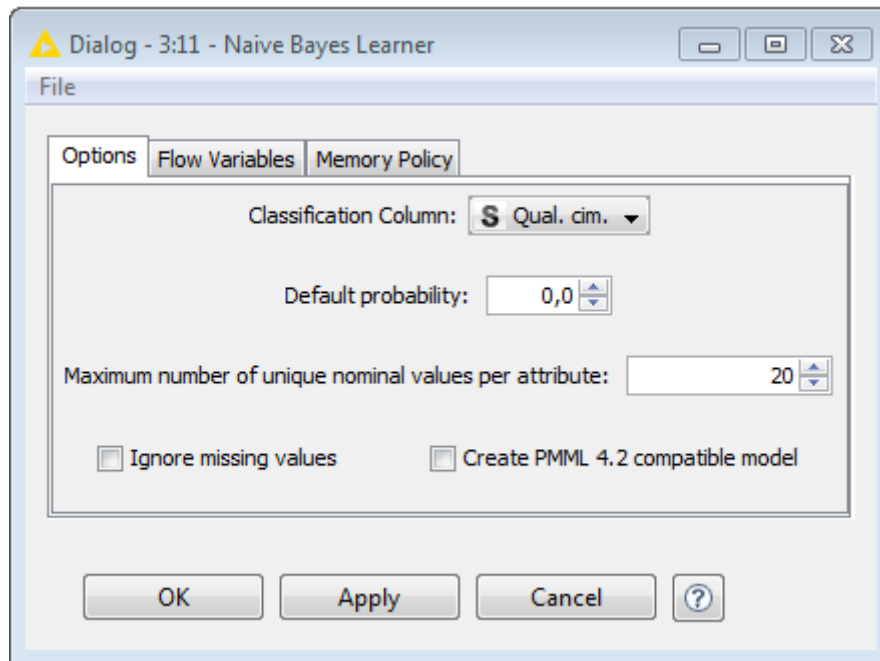


Figura 38 – Configuração do modelo de Naive Bayes

4.1.2 Árvores de Decisão

A árvore de decisão chega ao resultado da classificação pela execução de uma sequência de testes. Cada nó interno da árvore corresponde a um teste do valor de um dos atributos, e os ramos deste nó são identificados com os possíveis valores do teste. Cada nó folha da árvore está associada a uma classe e especifica o valor de retorno se aquela folha for alcançada. Cada um dos possíveis percursos da árvore, da raiz à folha, corresponde a uma regra de classificação. No espaço definido pelos atributos, cada folha corresponde a um hiper-cubo em que a interseção destes é vazia e a união é todo o espaço [14]. A representação da árvore de decisão e do espaço dos atributos é apresentada na Figura 39.

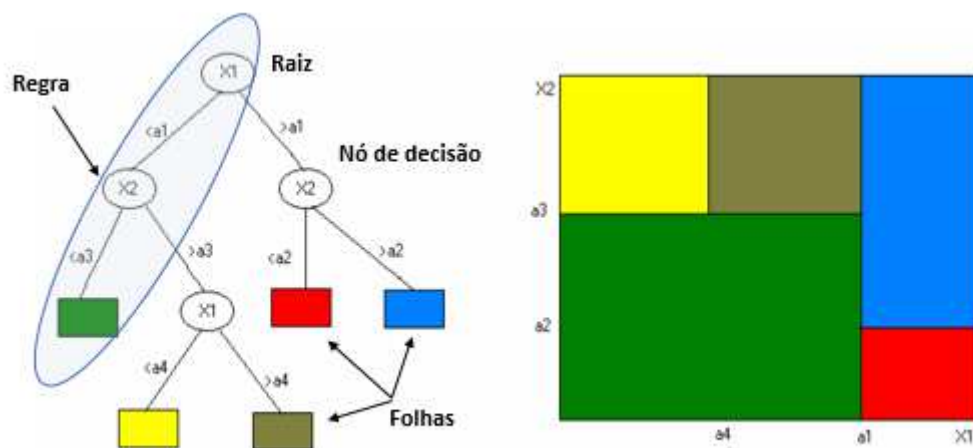


Figura 39 – Representação da árvore de decisão e particionamento do espaço de atributos [14]

A *workflow* gerado no *Knime* é apresentado na Figura 40. O algoritmo sempre particiona os registros em dois ramos quando o atributo testado é numérico. Para atributos nominais, o número de ramos será um valor entre dois e o número de valores possíveis para cada variável. [20]

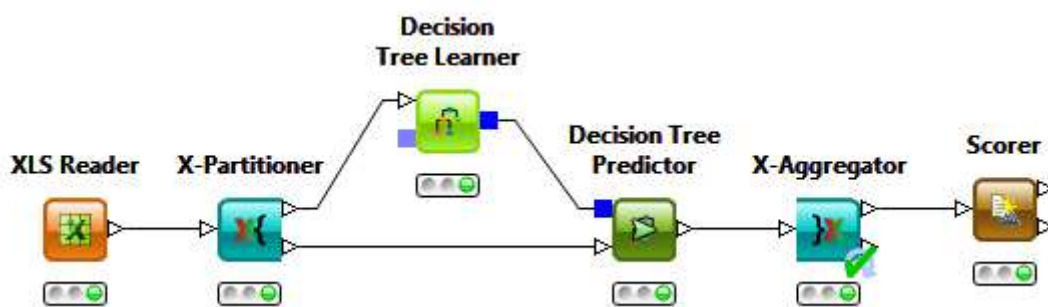


Figura 40 – Workflow Árvore de Decisão

A configuração do modelo consiste na definição dos seguintes parâmetros, conforme apresentado na Figura 41:

- Coluna que contém a variável alvo, ou seja, as classes (qualidade da cimentação);
- Índice de qualidade. Define qual o índice será utilizado para determinar o atributo para particionamento. A ideia principal é dividir os dados do nó-pai de forma a minimizar o grau de impureza dos nós-filhos, ou seja, fazer com que os nós-filhos contenham o maior desbalanceamento possível entre as classes. [29] São duas as opções disponíveis:
 - Gini Index: emprega um índice de dispersão estatístico proposto em 1912 [29];
 - Gain Ratio: o Ganho de Informação determina a medida de impureza pela comparação entre o grau de entropia do nó-pai com o grau de entropia dos nós-filhos. [29] A entropia representa a heterogeneidade do subconjunto em relação às classes. A razão de ganho aplica uma ponderação no ganho de informação [29];
- Método de poda. A poda reduz o tamanho da árvore, evitando que o modelo fique excessivamente ajustado ao conjunto de treinamento, podendo prejudicar a qualidade de predição. O método disponível no *Knime* é o MDL – Minimal Description Length. A outra opção disponível é desativar o processo de poda;
- Redução no erro de poda. Quando selecionado, o algoritmo utiliza um método de poda de pós-processamento. Iniciando das folhas, cada nó é substituído pela sua classe mais popular. O processo é realizado apenas caso não tenha decréscimo de acurácia. O objetivo é a simplificação da árvore e maior velocidade de processamento [20];
- Mínimo de registros por nó. Corresponde a um critério de parada. Caso o número de registros seja menor ou igual ao definido, o crescimento da árvore é interrompido [20];

- Ponto médio de particionamento. Quando selecionado, o particionamento para atributos numéricos irá ocorrer no ponto médio de seus valores. Caso não seja selecionado, o particionamento é determinado pelo ponto que maximiza o ganho de informação para aquele atributo.

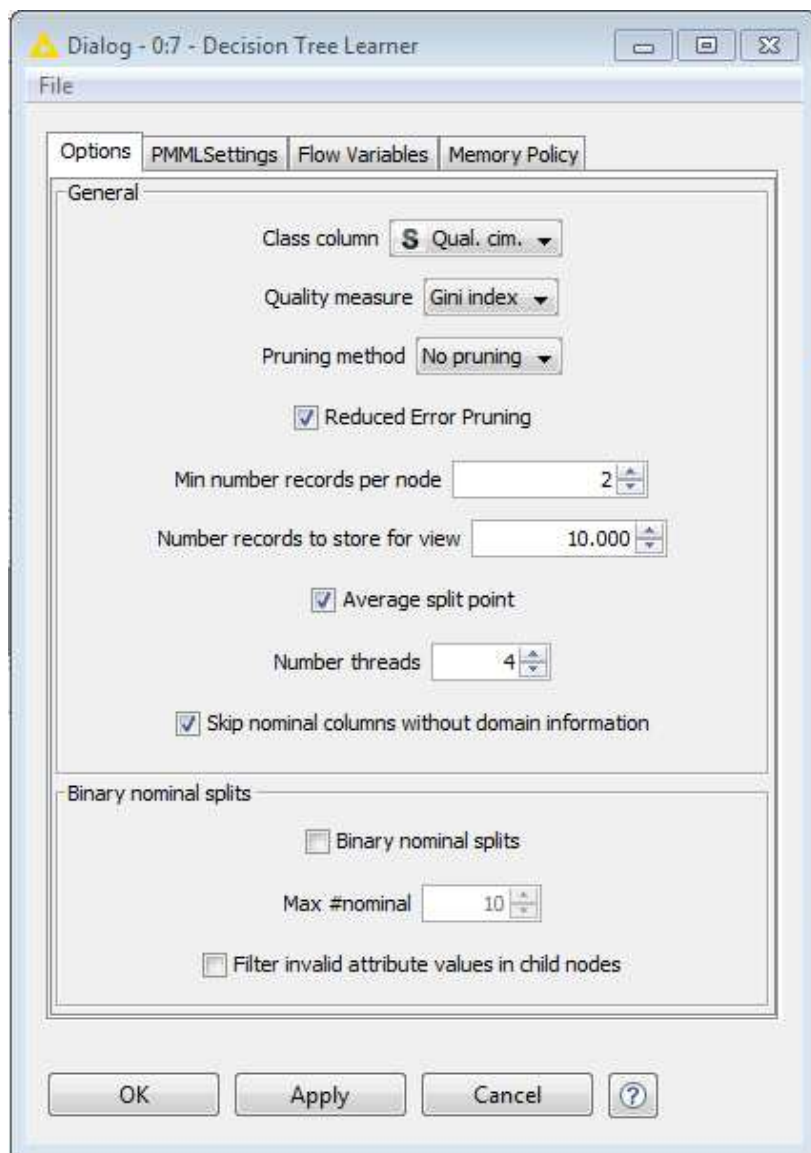


Figura 41 – Configuração do modelo de Árvore de Decisão

4.1.3 Tree Ensemble

Classificadores “*ensemble*” consistem em uma combinação de um conjunto de classificadores para a classificação de novas amostras. Geralmente, seu desempenho é melhor do que o desempenho dos classificadores individuais que o formaram.

O modelo Tree Ensemble é um tipo de classificador “ensemble” e combina então o resultado de várias Árvore de Decisão, retornando como predição final a classe predita com maior frequência.

O *workflow* gerado no *Knime* é apresentado na Figura 42. Cada modelo de árvore de decisão pode ser treinado em um conjunto diferente de registros e/ou de atributos. O modelo de predição irá utilizar a classe mais frequente. [20] Vários dos parâmetros disponíveis para configuração do modelo já foram discutidos na Seção 4.1.2, já que se referem à configuração das árvores de decisão. Os parâmetros relevantes, que ainda não foram discutidos, são listados a seguir e apresentados na Figura 43.

- Número de modelos. Indica o número de árvores de decisão a serem combinadas. Foram realizados alguns testes com os bancos de dados analisados, sem grandes diferenças nos resultados;
- Amostra de dados (Linhas). Possibilita o treinamento das diferentes árvores de decisão com frações do banco de dados. A opção foi desabilitada;
- Amostra de dados (Colunas). Possibilita o treinamento das diferentes árvores de decisão com diferentes atributos. Os resultados apresentados neste trabalho foram obtidos com o treinamento de todas as árvores de decisão contendo as informações de todos os atributos;
- Seleção de atributos. Possibilita que cada nó da árvore seja treinado com diferentes conjuntos de atributos. Os resultados apresentados neste trabalho foram obtidos com o treinamento de todos os nós contendo as informações de todos os atributos.

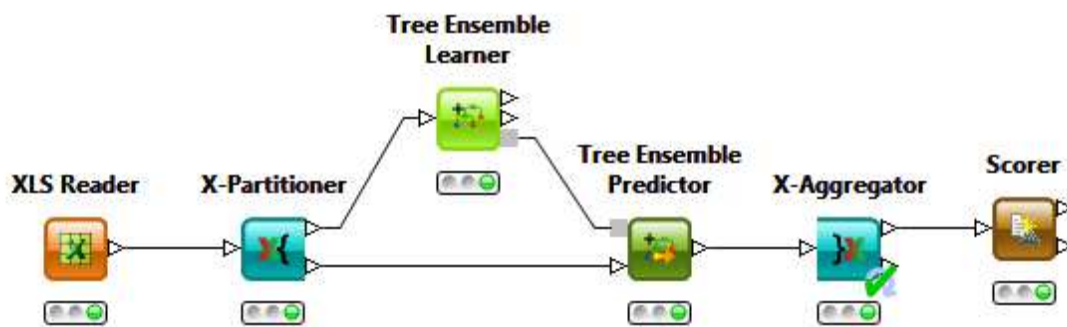


Figura 42 – Workflow Tree Ensemble

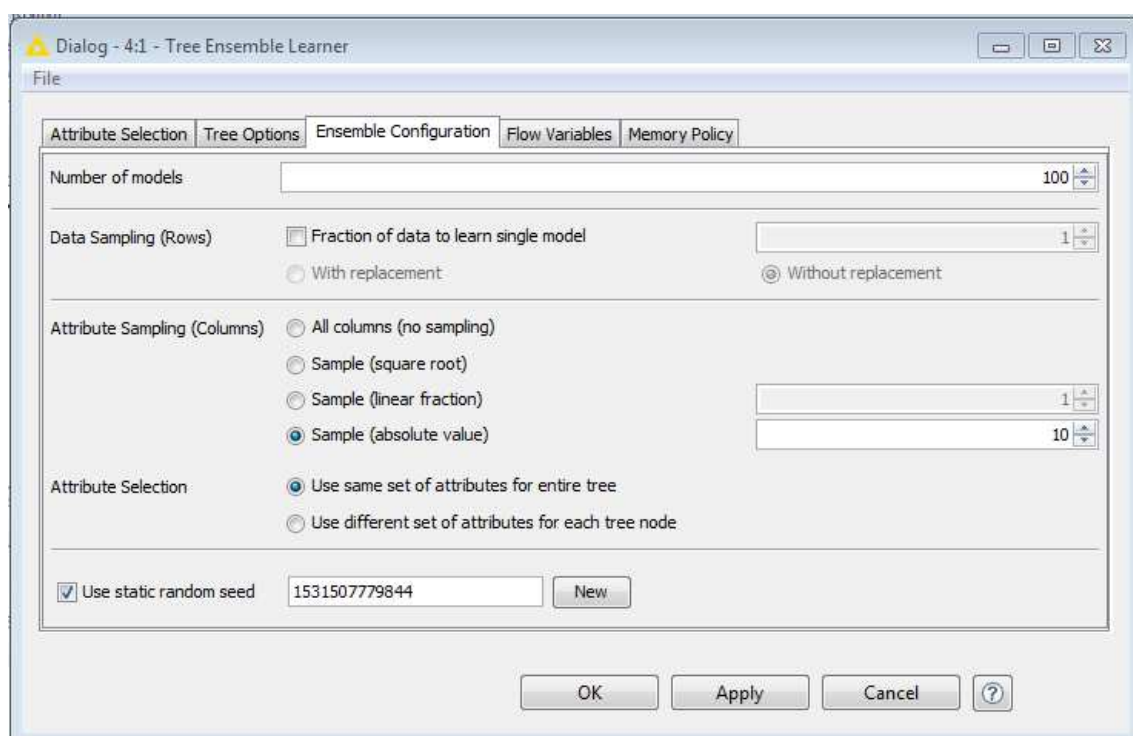


Figura 43 – Configuração do modelo Tree Ensemble

4.1.4 Redes Neurais

Algoritmos de redes neurais artificiais são inspirados no funcionamento do cérebro humano e nas redes neurais biológicas. As redes neurais artificiais contém unidades de processamento (neurônios artificiais) conectadas entre elas por canais de comunicação que estão associados a determinado peso.

A Figura 44 apresenta o funcionamento de um neurônio artificial, proposto inicialmente por MacCulloch e Pitts e aprimorado por Rosenblatt. Cada sinal de entrada (x_m) é multiplicado pelo peso (w_{kp}), que indica sua influência no sinal de saída do neurônio, e são então somados. O resultado desse somatório será a entrada de uma função de ativação, que restringe a amplitude da saída de um neurônio, limitando o intervalo de saída para um valor finito. O neurônio possui ainda uma entrada externa, conhecida como “bias”, a qual tem o efeito de aumentar ou diminuir a entrada da função de ativação.

Os pesos serão determinados durante a etapa de treinamento do modelo, em que os algoritmos farão os ajustes necessários a partir de valores de entrada e saída previamente conhecidos.

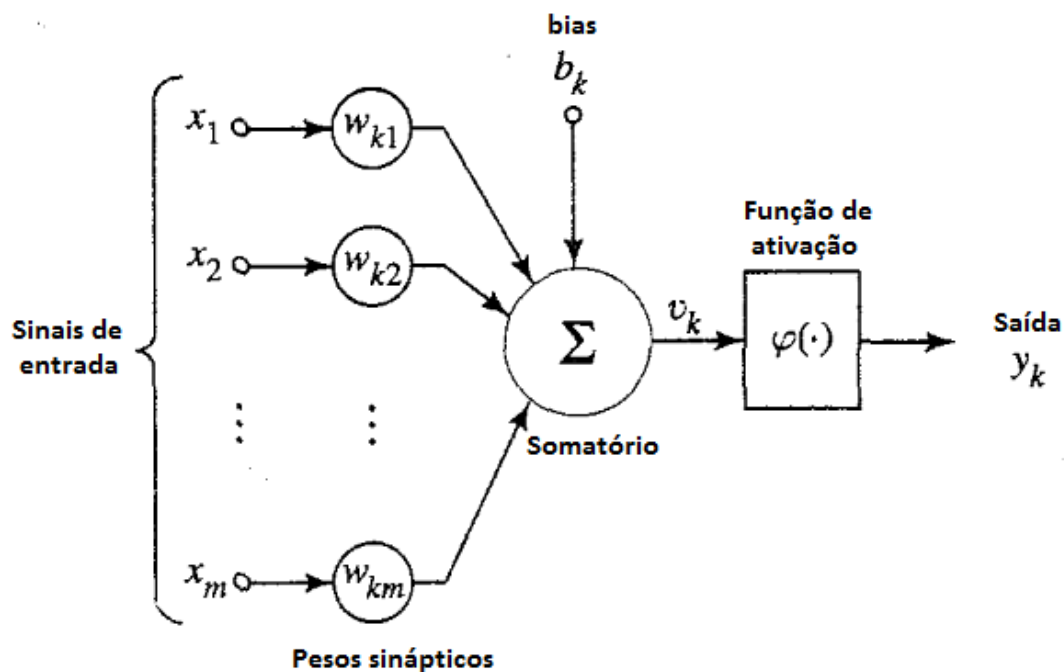


Figura 44 – Funcionamento de um neurônio artificial (Traduzido de [16])

Uma das arquiteturas mais conhecidas de redes neurais é a Perceptron de multi-camadas (MLP – *Multi-layer Perceptron*), a qual é formada por uma camada de entrada, uma ou mais camadas ocultas (intermediárias) e uma camada de saída. Os neurônios de cada uma das camadas ocultas e da camada de saída irá receber como dado de entrada a saída da camada anterior. A Figura 45 apresenta um exemplo de rede MLP com 4

neurônios na camada de entrada, uma camada oculta com 3 neurônios e uma camada de saída com 2 neurônios. [16]

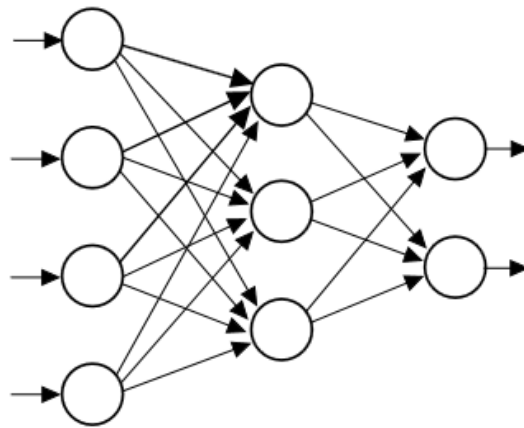


Figura 45 – Exemplo de rede MLP [16]

A implementação do modelo no *Knime* é apresentada na Figura 46. A configuração do modelo consiste na definição dos seguintes parâmetros, conforme apresentado na Figura 47:

- Número máximo de iterações durante o treinamento;
- Número de camadas intermediárias. Foram realizadas análise com diferentes configurações, conforme descrito na Seção 5;
- Número de neurônios por camada. Foram realizadas análise com diferentes configurações, conforme descrito na Seção 5;
- Coluna que contém a variável alvo, ou seja, as classes (qualidade da cimentação).

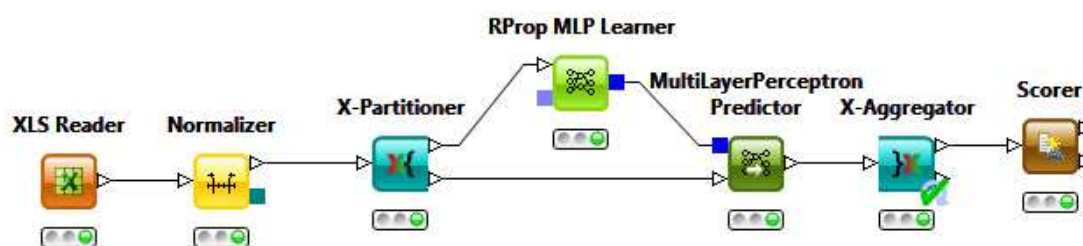


Figura 46 – Workflow Redes Neurais

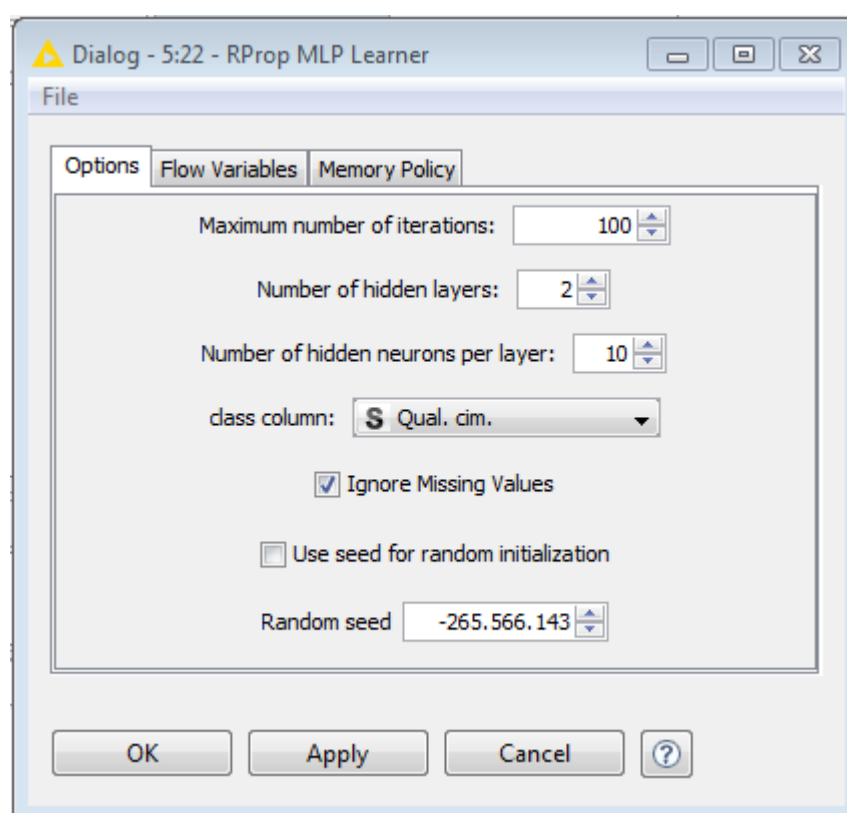


Figura 47 – Configuração do modelo de Redes Neurais

4.2 Validação dos Modelos

As estatísticas dos modelos estudados foram calculadas em validação cruzada de forma a avaliar a capacidade de generalização dos mesmos. Cada uma das três bases de dados foi dividida em 10 subconjuntos de mesmo tamanho, mantendo a proporção entre

as classes, dentre os quais 9 são utilizados para treinamento do modelo e 1 utilizado para teste. O processo é repetido 10 vezes, cada uma delas com um subconjunto de teste diferente, e analisada então a média das métricas de avaliação de desempenho.

As métricas de avaliação utilizadas nesse trabalho são calculadas a partir da matriz de confusão, a qual apresenta o número de classificações preditas de maneira correta e incorreta para cada classe. Cada linha da matriz representa a classe real das amostras e cada coluna a classe predita pelo classificador, conforme *Tabela 4*.

Tabela 4 - Exemplo de Matriz de Confusão (Adaptado de [12])

	$\hat{C}_1$ (predita)	$\hat{C}_2$ (predita)
C_1 (real)	Verdadeiros Positivos (Registros da classe C_1 corretamente classificados pelo modelo)	Falsos Negativos (Registros da classe C_1 incorretamente classificados pelo modelo)
C_2 (real)	Falsos Positivos (Registros da classe C_2 incorretamente classificados pelo modelo)	Verdadeiros Negativos (Registros da classe C_2 corretamente classificados pelo modelo)

Como estamos tratando de bases de dados desbalanceadas, conforme apresentado na Seção 3 pelos gráficos de divisão dos dados entre as classes, e nossa maior preocupação é com a classe minoritária, buscamos métricas de avaliação de desempenho dos modelos que levem em consideração o erro para cada classe. Dessa forma, foram selecionadas as métricas de erro listadas a seguir. Os valores calculados para as métricas apresentados nas seções seguintes foram obtidos para a classe negativa, não obtenção de isolamento hidráulico, já que essa é a nossa classe prioritária.

4.2.1 Precisão

A medida de precisão tem como objetivo identificar qual a proporção de amostras classificadas corretamente dentre as classificações realizadas para uma

determinada classe. Considerando C_1 a classe positiva (para este trabalho, isolamentos obtidos) e C_2 a classe negativa (falha de isolamento), a métrica é definida para cada uma das classes, portanto, conforme Equações 3 e 4, respectivamente.

$$PRE(C_1) = \frac{VP}{VP + FP} \quad (3)$$

Onde $PRE(C_1)$ = Precisão da classe positiva

VP = Verdadeiros Positivos

FP = Falsos Positivos

$$PRE(C_2) = \frac{VN}{VN + FN} \quad (4)$$

Onde $PRE(C_2)$ = Precisão da classe negativa

VN = Verdadeiros Negativos

FN = Falsos Negativos

Para o problema analisado, a medida de precisão está relacionada ao fator “economia”. Quanto maior a precisão da classe negativa, relacionada à falha de isolamento hidráulico, menor o número de falsos negativos, ou seja, menor o número de registros com boa qualidade da cimentação preditos pelo modelo como falha. Aumentando então a precisão da classe negativa, permite-se a opção pela supressão da operação de perfilagem dos poços com boa qualidade da cimentação.

4.2.2 Recuperação

A medida de recuperação identifica a frequência em que as amostras de uma determinada classe são classificadas corretamente, ou seja, a proporção entre o número de exemplos classificados corretamente para uma determinada classe e a quantidade total de exemplos que pertencem a essa classe. Dessa forma, temos a recuperação para a

classe positiva e para a classe negativa definidas segundo as Equações 5 e 6, respectivamente.

$$REC(C_1) = \frac{VP}{VP + FN} \quad (5)$$

Onde $REC(C_1)$ = Recuperação da classe positiva

VP = Verdadeiros Positivos

FN = Falsos Negativos

$$REC(C_2) = \frac{VN}{VN + FP} \quad (6)$$

Onde $REC(C_2)$ = Recuperação da classe negativa

VN = Verdadeiros Negativos

FP = Falsos Positivos

A medida de recuperação está relacionada ao fator “segurança” do problema analisado. Um aumento da recuperação da classe negativa, relacionada à falha de isolamento hidráulico, indica um menor número de falsos positivos, ou seja, menor o número de registros com falha de cimentação preditos pelo modelo como sucesso na obtenção de isolamento hidráulico. O fator de 100% de recuperação da classe negativa, garante, portanto, que todos os registros com falha de cimentação estão sendo identificados para confirmação por meio de perfis acústicos. Dessa forma, é a métrica com maior relevância para as análises realizadas, buscando-se um aumento na precisão da classe negativa apenas após a conquista de alta taxa de recuperação da classe negativa.

4.2.3 Medida F

A medida F consiste na média ponderada das medidas de precisão (PRE) e recuperação (REC), conforme Equação 7.

$$F(C_i) = \frac{2 PRE(C_i) \times REC(C_i)}{PRE(C_i) + REC(C_i)} \quad (7)$$

Combinando essas duas métricas, a medida F propõe um valor único que indique a qualidade geral do modelo, de maneira que, quanto maior o valor de F (mais próximo de 1), melhor o desempenho em termos de precisão e recuperação simultaneamente.

4.2.4 AUC

A análise ROC (do inglês, *Receiver Operating Characteristic*) é um método gráfico para avaliação de modelos de predição. O gráfico ROC, apresentado na Figura 48, é construído com a taxa de verdadeiros positivos no eixo das ordenadas e com a taxa de falsos positivos no eixo das abscissas. [25]

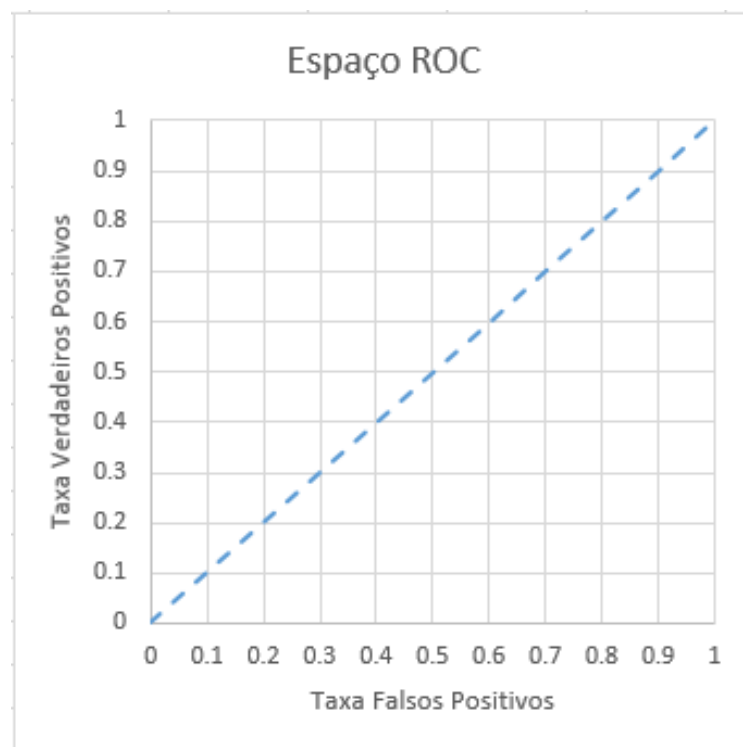


Figura 48 - Espaço ROC

O espaço ROC mostra o desempenho do classificador relacionando o quanto o classificador acerta da classe positiva e o quanto deixa de errar da classe negativa. O modelo perfeito é o ponto (0,1) do gráfico, que representa aquele em que o modelo prevê corretamente todos os exemplos de ambas as classes. Já o ponto (1,0) representa o modelo que faz todas as previsões erradas. Os modelos localizados sobre a diagonal do espaço ROC têm desempenho equivalente a um modelo aleatório. [12][25]

Uma métrica obtida a partir do espaço ROC é a área sobre a curva, AUC (do inglês, *Area Under Curve*). Para definir a área a ser calculada, liga-se as coordenadas do classificador aos pontos (0,0) e (1,1), formando um quadrilátero, conforme Figura 49. Simplificadamente, o valor de AUC é calculado pela média entre a taxa de verdadeiro positivo e verdadeiro negativo, segundo Equação 8.

$$AUC = \frac{REC(C_1) + REC(C_2)}{2} \quad (8)$$

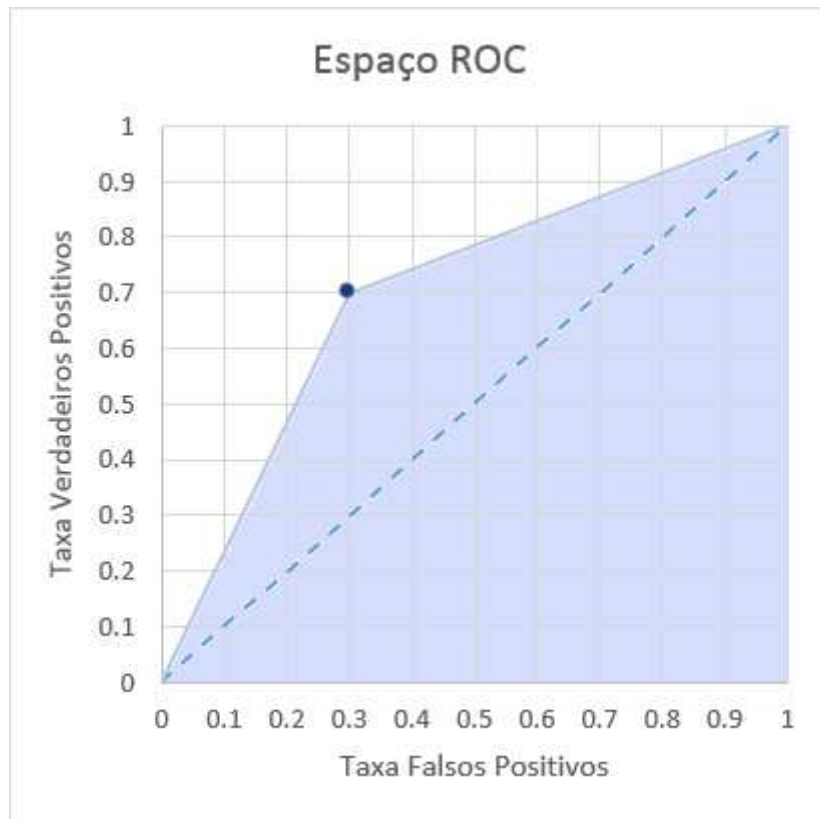


Figura 49 – AUC

Para problemas de múltiplas classes, são desenhadas curvas ROC para cada uma das classes em relação às demais sendo a área abaixo da curva calculada separadamente para cada uma das classes. O valor de AUC final para o classificador, será calculado conforme Equação 9, a média do AUC de cada classe, ponderada pela estimativa de propriedade a priori da classe correspondente [12]. Uma desvantagem dessa abordagem é a sensibilidade ao desbalanceamento entre as classes, já que a probabilidade a priori é considerada.

$$AUC = \sum_{i=1}^M AUC(C_i)P(C_i) \quad (9)$$

5 Análise dos Resultados

A medida de validação considerada de maior relevância para as análises realizadas é a recuperação da classe “Ruim”, já que o objetivo deste trabalho é a supressão das operações de perfilagem para avaliação da qualidade da cimentação, mas de maneira a não prejudicar a segurança do poço. Dessa forma, as cimentações com potencial de falha na cimentação devem continuar sendo perfiladas, para que se possa ter mais informações à cerca da segurança do poço e, em casos em que haja necessidade, corrigir a cimentação do poço. A medida de recuperação da classe “Ruim” permitirá então a identificação da frequência em que as amostras dessa classe são classificadas corretamente, de maneira que, quanto melhor seu resultado, maior a confiabilidade para a decisão de não perfilar os demais poços, que não foram acusados pelo modelo como potencialmente falhos.

Os modelos de classificação Naive Bayes, Árvores de Decisão, Tree Ensemble e Perceptron de Múltiplas Camadas foram testados para BD1, BD2 e BD3. Para o modelo de Árvore de Decisão, testou-se as possíveis combinações de configuração, variando os parâmetros Medida de Qualidade, Método de Poda e Ponto Médio de Particionamento, conforme descrito na Seção 4.1.2. Para as análises em BD2, não faz sentido a variação no parâmetro que indica o método para determinação do ponto de particionamento pois ele só é válido para variáveis numéricas e, em BD2, todas as variáveis são binárias. As análises foram realizadas variando então os parâmetros “Medida de Qualidade” e “Método de Poda”. Para o modelo de Redes Neurais, foram testadas configurações variando o número de camadas, 1 e 2 camadas, e a quantidade de neurônios, 10 e 20 neurônios por camada. A Tabela 5 contém um resumo do número de configurações testadas de cada modelo para cada um dos Bancos de Dados analisados, totalizando 76 testes realizados.

Tabela 5 – Resumo da quantidade de modelos de classificação testados

	BD1		BD2		BD3	
	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes
Naive Bayes	1	1	1	1	1	1
Árvores de Decisão	8	8	4	4	8	8
Tree Ensemble	1	1	1	1	1	1
Redes Neurais	4	4	4	4	4	4

5.1 Banco de Dados 1

Em um primeiro momento, foram testados os modelos de classificação para BD1.

A Tabela 6 apresenta os resultados para o banco de dados com 2 classes e a Tabela 7 com 4 classes para as possíveis variações do modelo de Árvore de Decisão. Para o problema com 2 classes, os piores resultados foram encontrados para as situações em que se utilizou conjuntamente o Gini Index e o método de poda MDL, apresentando precisão e recuperação da classe “Ruim” nulas. A diferença no resultado obtido para as demais configurações não é significativa e em nenhuma delas é possível ter confiança na predição pelo modelo. Para o problema com 4 classes, foram 5 dentre as 8 combinações dos parâmetros possíveis que apresentaram resultados de precisão e recuperação da classe “Ruim” nulas, mostrando uma maior tendência de classificação como “Excelente”, a classe majoritária. Dentre as outras 3 configurações, a que apresentou melhor desempenho foi a que utiliza o Gini Index, não realiza processo de poda e é utilizado o máximo ganho para determinar o ponto de particionamento das variáveis numéricas.

Tabela 6 – Comparação de desempenho de classificação de BD1 com 2 classes com diferentes configurações do modelo de Árvore de Decisão

	Gini Index				Gain Ratio			
	Sem poda		MDL		Sem poda		MDL	
	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho
Prec. "Ruim"	5,9%	11,8%	0%	0%	9,1%	7,7%	40,0%	33,3%
Rec. "Ruim"	4,2%	8,0%	0%	0%	3,8%	3,8%	7,7%	3,8%
Medida F	4,9%	9,5%	0%	0%	5,4%	5,1%	12,9%	6,9%
AUC	50,2%	52,3%	49,7%	49,5%	50,8%	50,5%	53,5%	51,7%

Tabela 7 - Comparação de desempenho de classificação de BD1 com 4 classes com diferentes configurações do modelo de Árvore de Decisão

	Gini Index				Gain Ratio			
	Sem poda		MDL		Sem poda		MDL	
	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho
Prec. "Ruim"	14,3%	20,0%	20,0%	0%	0%	0%	0%	0%
Rec. "Ruim"	8,0%	16,0%	3,8%	0%	0%	0%	0%	0%
Medida F	10,3%	17,8%	6,5%	0%	0%	0%	0%	0%
AUC	73,4%	73,6%	80,8%	83,8%	90,5%	90,0%	92,8%	92,8%

Para as variações de Redes Neurais testadas os resultados são apresentados na Tabela 8 e na Tabela 9 para os problemas tratados com 2 classes e 4 classes, respectivamente. Para o problema simplificado, com 2 classes, as configurações com 1 camada e 10 neurônios, 2 camadas com 10 neurônios e 2 camadas com 20 neurônios, obtiveram o mesmo desempenho considerando a recuperação da classe negativa, falha de isolamento hidráulico. Sendo assim, foi escolhido prosseguir com as análises considerando a rede de 2 camadas e 20 neurônios, por ser a de maior precisão da classe negativa. Para o problema com 4 classes, observaram-se também melhores resultados para a rede com 2 camadas com 10 neurônios em cada uma delas, neste caso, já na etapa de comparação da recuperação da classe negativa.

Tabela 8 - Desempenho da classificação de BD1 com 2 classes por Redes Neurais

	1 camada com 10 neurônios	1 camada com 20 neurônios	2 camadas com 10 neurônios	2 camadas com 20 neurônios
Prec. "Ruim"	16,7%	11,1%	18,8%	15,8%
Rec. "Ruim"	11,5%	7,7%	11,5%	11,5%
Medida F	13,6%	9,1%	14,3%	13,3%
AUC	54,0%	52,0%	54,3%	53,9%

Tabela 9 - Desempenho da classificação de BD1 com 4 classes por Redes Neurais

	1 camada com 10 neurônios	1 camada com 20 neurônios	2 camadas com 10 neurônios	2 camadas com 20 neurônios
Prec. "Ruim"	15,4%	20,0%	25%	27,3%
Rec. "Ruim"	7,7%	11,5%	19,2%	11,5%
Medida F	10,3%	14,6%	21,7%	16,2%
AUC	85,8%	81,8%	81,2%	81,7%

A Tabela 10 apresenta o desempenho dos modelos de classificação. Considerando principalmente as informações de recuperação da classe negativa, que mede a proporção de acertos de predição de registros com falha na cimentação, o modelo que teve melhores resultados foi o Naive Bayes, apresentando desempenho similar para o problema com 2 classes e com 4 classes. No entanto, esse resultado é muito longe do esperado, e não fornece confiabilidade para a não execução de perfilagem para avaliação da qualidade da cimentação com base nessas predições, não sendo, portanto, satisfatório.

Tabela 10 - Desempenho da classificação de BD1

	Naive Bayes		Árvores de Decisão		Tree Ensemble		MLP	
	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes
Prec. "Ruim"	13,6%	14,3%	40,0%	20,0%	12,1%	8,1%	18,8%	25%
Rec. "Ruim"	23,1%	23,1%	7,7%	16,0%	15,4%	11,5%	11,5%	19,2%
Medida F	17,1%	17,6%	12,9%	17,8%	13,6%	9,5%	14,3%	21,7%
AUC	57,2%	65,0%	53,5%	73,6%	54,4%	72,2%	54,3%	81,2%

5.2 Banco de Dados 2

Utilizou-se a mesma metodologia desenvolvida para a análise de BD1 para avaliação de desempenho dos modelos aplicados ao Banco de Dados 2.

A Tabela 11 e a Tabela 12 apresentam, respectivamente, o resultado de comparação do desempenho das diferentes configurações do modelo de árvore de decisão para BD2 com 2 e 4 classes. Para o problema de 2 classes, quando utilizados conjuntamente o *Gain Ratio* e o método de poda, o modelo apresentou tendência de classificar todos os registros como "Isolamento Obtido", obtendo resultados de precisão e recuperação da classe "Ruim" nulos. As outras configurações apresentaram resultados muito similares. Para o problema de 4 classes, quando utilizados conjuntamente o *Gain Ratio* e o método de poda, apesar dos resultados não terem sido nulos como para o problema de 2 classes, foram muito inferiores aos demais, que apresentaram desempenho similares.

Tabela 11 – Comparação de desempenho de classificação de BD2 com 2 classes com diferentes configurações do modelo de Árvore de Decisão

	Gini Index		Gain Ratio	
	Sem poda	MDL	Sem poda	MDL
Prec. "Ruim"	50,0%	53,3%	47,1%	0%
Rec. "Ruim"	26,7%	26,7%	26,7%	0%
Medida F	34,8%	35,6%	34%	0%
AUC	61,5%	61,8%	61,3%	50,0%

Tabela 12 - Comparação de desempenho de classificação de BD2 com 4 classes com diferentes configurações do modelo de Árvore de Decisão

	Gini Index		Gain Ratio	
	Sem poda	MDL	Sem poda	MDL
Prec. "Ruim"	50,0%	55,0%	44,8%	33,3%
Rec. "Ruim"	43,3%	36,7%	43,3%	6,7%
Medida F	46,4%	44,0%	44,1%	11,1%
AUC	93,9%	94,0%	94,2%	88,6%

Assim como nas análises das configurações de rede MLP para o problema de 2 classes de BD1, não foram observadas variações significativas de desempenho para o problema de 2 classes de BD2, conforme Tabela 13. Dessa forma, foi escolhido novamente trabalhar com a rede de 1 camada com 10 neurônios devido ao menor custo

computacional. Para o problema de 4 classes, a Tabela 14 apresenta melhores resultados para a rede com 2 camadas com 20 neurônios em cada uma delas.

Tabela 13 - Desempenho da classificação de BD2 com 2 classes por Redes Neurais

	1 camada com 10 neurônios	1 camada com 20 neurônios	2 camadas com 10 neurônios	2 camadas com 20 neurônios
Prec. "Ruim"	52,4%	47,4%	50,0%	52,6%
Rec. "Ruim"	36,7%	30,0%	30,0%	33,3%
Medida F	43,1%	37,7%	37,5%	40,8%
AUC	66,0%	62,7%	60,4%	64,6%

Tabela 14 - Desempenho da classificação de BD2 com 4 classes por Redes Neurais

	1 camada com 10 neurônios	1 camada com 20 neurônios	2 camadas com 10 neurônios	2 camadas com 20 neurônios
Prec. "Ruim"	55,0%	45,0%	47,4%	65,0%
Rec. "Ruim"	36,7%	30,0%	30,0%	43,3%
Medida F	44,0%	36,0%	36,7%	52,0%
AUC	93,3%	91,0%	91,7%	94,4%

A comparação entre o desempenho dos quatro modelos aplicados a BD2 é apresentada na Tabela 15, na qual se observa que, assim como na análise de BD1, o modelo com melhores resultados é o Naive Bayes. No entanto, para BD2, houve diferença expressiva entre o desempenho do modelo para o problema de 2 classes e de 4

classes, com melhores resultados para o último, chegando a 70% de recuperação para a classe negativa.

Tabela 15 - Desempenho da classificação de BD2

	Naive Bayes		Árvores de Decisão		Tree Ensemble		MLP	
	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes
Prec. "Ruim"	23,9%	25,3%	50,0%	50,0%	45,8%	45,5%	52,4%	65,0%
Rec. "Ruim"	56,7%	70,0%	26,7%	43,3%	36,7%	33,3%	36,7%	43,3%
Medida F	33,7%	37,2%	34,8%	46,4%	40,7%	38,5%	43,1%	52,0%
AUC	65,9%	79,8%	61,5%	93,9%	65,4%	93,7%	66,0%	94,4%

5.3 Banco de Dados 3

As últimas análises dos modelos de classificação foram para aplicação no BD3.

O desempenho dos classificadores com as diferentes configurações de árvore de decisão para BD3 de 2 e 4 classes é apresentado, respectivamente, na Tabela 16 e na Tabela 17. Para as análises com 2 e com 4 classes, quando utilizados conjuntamente Gain Ratio e o método de poda MDL, novamente, foram obtidos resultados nulos, sendo todos os registros classificados como "Isolamento Obtido". Dentre as demais classificações, não há diferenças significativas de desempenho.

Tabela 16 – Comparação de desempenho de classificação de BD3 com 2 classes com diferentes configurações do modelo de Árvore de Decisão

	Gini Index				Gain Ratio			
	Sem poda		MDL		Sem poda		MDL	
	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho
Prec. "Ruim"	34,5%	38,1%	31,6%	40,0%	29,0%	28,0%	0%	0%
Rec. "Ruim"	35,7%	28,6%	21,4%	28,6%	32,1%	25,0%	0%	0%
Medida F	35,1%	32,7%	25,5%	33,3%	30,5%	26,4%	0%	0%
AUC	63,4%	61,3%	57,7%	61,5%	60,9%	58,3%	56,7%	56,7%

Tabela 17 - Comparação de desempenho de classificação de BD3 com 4 classes com diferentes configurações do modelo de Árvore de Decisão

	Gini Index				Gain Ratio			
	Sem poda		MDL		Sem poda		MDL	
	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho	Ponto Médio	Máx Ganho
Prec. "Ruim"	32,0%	29,6%	26,5%	26,9%	31,2%	22,2%	0%	0%
Rec. "Ruim"	28,6%	28,6%	32,1%	25,0%	35,7%	28,6%	0%	0%
Medida F	30,2%	29,1%	29,0%	25,9%	33,3%	25,0%	0%	0%
AUC	80,6%	78,1%	89,6%	89,7%	77,8%	81,8%	88,4%	88,0%

A Tabela 18 indica que foram obtidos desempenhos similares para as configurações de redes neurais com 1 camada com 20 neurônios, 2 camadas com 10 neurônios e 2 camadas com 20 neurônios, considerando a recuperação da classe negativa para o BD3 3 com 2 classes. Analisando, portanto, a precisão da classe negativa, o melhor desempenho foi obtido utilizando-se 2 camadas com 20 neurônios em cada uma delas. A Tabela 19 apresenta os resultados de desempenho para o problema de 4 classes, com resultados similares de recuperação da classe “Ruim” para 1 camada com 10 neurônios e com 20 neurônios, sendo o desempenho superior para a configuração com 20 neurônios, considerando a precisão da classe “Ruim”.

Tabela 18 - Desempenho da classificação de BD3 com 2 classes por Redes Neurais

	1 camada com 10 neurônios	1 camada com 20 neurônios	2 camadas com 10 neurônios	2 camadas com 20 neurônios
Prec. "Ruim"	45,0%	47,8%	40,7%	50,0%
Rec. "Ruim"	32,1%	39,3%	39,3%	39,3%
Medida F	37,5%	43,1%	40,0%	44,0%
AUC	63,4%	66,8%	65,9%	67,0%

Tabela 19 - Desempenho da classificação de BD3 com 4 classes por Redes Neurais

	1 camada com 10 neurônios	1 camada com 20 neurônios	2 camadas com 10 neurônios	2 camadas com 20 neurônios
Prec. "Ruim"	54,2%	56,5%	47,8%	60,0%
Rec. "Ruim"	46,4%	46,4%	39,3%	42,9%
Medida F	50,0%	51,0%	43,1%	50,0%
AUC	86,8%	87,7%	85,0%	87,0%

Assim como para os demais bancos de dados analisados, o modelo com melhor desempenho para predição utilizando as informações presentes em BD3 é o Naive Bayes, conforme Tabela 20. Para BD3 não há diferenças significativas entre a análise do problema simplificado, com 2 classes, e do problema original, com 4 classes.

Tabela 20 - Desempenho da classificação de BD3

	Naive Bayes		Árvores de Decisão		Tree Ensemble		MLP	
	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes
Prec. "Ruim"	32,2%	33,9%	34,5%	32,0%	25,0%	40,7%	50,0%	56,5%
Rec. "Ruim"	67,9%	67,9%	35,7%	28,6%	21,4%	39,3%	39,3%	46,4%
Medida F	43,7%	45,2%	35,1%	30,2%	23,1%	40,0%	44,0%	51,0%
AUC	74,5%	74,4%	63,4%	80,6%	56,4%	80,3%	67,0%	87,7%

5.4 Comparação entre os resultados dos três bancos de dados

Para comparar o resultado dos modelos de classificação quando aplicados a cada um dos três bancos de dados disponíveis, são apresentadas na Tabela 21 as métricas de desempenho do modelo Naive Bayes de cada um deles, já que foi o algoritmo que apresentou maior eficiência para todos eles, considerando principalmente a recuperação da classe “Ruim”, conforme já discutido.

O modelo de Naive Bayes surpreendeu por ser o mais eficiente, já que considera a propriedade a priori e os bancos de dados analisados têm grande desbalanceamento. No entanto, os demais modelos se mostraram menos eficientes para o problema analisado.

Quando considerado o problema com 4 classes, observa-se que os resultados foram similares para BD2 e BD3, sendo essas muito melhores que os resultados obtidos para BD1. Para o problema com 2 classes, apenas o resultado do modelo aplicado a BD3 tem resultado similar àqueles identificados como os melhores para o problema de 4 classes. Os resultados para 2 classes em BD1 e BD2 são inferiores.

Devido à alta criticidade da decisão pela não operação de perfilagem relacionada ao potencial de ocorrência de um acidente em casos da não obtenção de isolamentos hidráulicos requeridos, conforme discutido na Seção 2.1, deseja-se que a recuperação da classe negativa, seja o mais próximo possível de 100%, diminuindo assim o risco operacional. Dessa forma nenhum dos resultados provê confiabilidade para a decisão de suprimir a operação de perfilagem para avaliar a qualidade da cimentação.

Tabela 21 - Desempenho da classificação comparativo para os três bancos de dados utilizando o modelo Naive Bayes

	BD1		BD2		BD3	
	2 classes	4 classes	2 classes	4 classes	2 classes	4 classes
Prec. "Ruim"	13,6%	14,3%	23,9%	25,3%	32,2%	33,9%
Rec. "Ruim"	23,1%	23,1%	56,7%	70,0%	67,9%	67,9%
Medida F	17,1%	17,6%	33,7%	37,2%	43,7%	45,2%
AUC	57,2%	65,0%	65,9%	79,8%	74,5%	74,4%

6 Conclusões

Este trabalho inicia um projeto de pesquisa de relação entre as fases de planejamento, operação e o resultado final da qualidade da cimentação de poços de petróleo offshore com base em um histórico de registros de perfilagem acústica para avaliação da cimentação. A identificação da combinação de parâmetros que possibilitem a falha de obtenção de isolamento hidráulico requerido pode permitir que a perfilagem de verificação da cimentação seja desnecessária em alguns casos, sendo possível atestar a qualidade da cimentação apenas pela análise desses parâmetros.

Para as análises realizadas foram utilizados registros presentes em três bancos de dados distintos, a fim de se verificar os parâmetros operacionais mais representativos para o problema.

Foi observado que o fator operacional que mais afeta a qualidade da cimentação é a ocorrência de perdas de circulação, principalmente durante o bombeio da pasta de cimento. Cenários de perda podem ter consequências amenizadas quando se tem conhecimento dessa possibilidade, de forma a utilizar sistemas de pasta e colchão especiais e programar um excesso de pasta a ser bombeado compatível ao cenário esperado.

Um bom ponto de partida para a decisão sobre a necessidade ou não da corrida de perfis acústicos para atestado da qualidade da cimentação e inferência dos isolamentos hidráulicos requeridos é a análise dos parâmetros operacionais selecionados e listados no APÊNDICE B, sempre levando em consideração o conhecimento do cenário, como a expertise da operadora em cimentar aquelas formações e os resultados anteriores com o mesmo sistema de pasta.

Dentre os modelos de classificação utilizados, foi uma surpresa constatar que o mais simples dentre eles, Naive Bayes, obteve melhores resultados para todos os três bancos de dados. O algoritmo, quando aplicado aos bancos de dados 2 e 3 com 4 classes e ao banco de dados 3 com 2 classes, apresentou melhor resultado de predição, levando-se em consideração, principalmente, a recuperação da classe minoritária já que esse é o método de validação que melhor se aplica à confiabilidade na supressão da operação de

perfilagem para avaliar a qualidade da cimentação, sem prejudicar a condição de segurança do poço.

Acredita-se que os melhores resultados conquistados com os modelos aplicados aos bancos de dados 2 e 3 se devem à seleção das variáveis representativas das características da operação de cimentação, enquanto que o banco de dados 1 contém a maior parte das variáveis relacionadas às condições de poço e planejamento da cimentação. Devido à maturidade da operadora em cimentar nos cenários estudados e a grande preocupação com o planejamento das operações, realizando todos os testes e simulações adequadas, são selecionadas as características da pasta de cimento que melhor se enquadram ao cenário ao qual será submetida. Dessa forma, os resultados mostram que a companhia está no caminho certo quanto à fase de planejamento da cimentação, não sendo estes fatores determinantes para a falha de isolamento.

De maneira geral, as análises considerando o problema com 4 classes (Excelente, Boa, Média e Ruim) apresentaram melhores resultados de predição quando comparadas às análises considerando o problema com 2 classes (Isolamento Obtido e Falha de Isolamento). Esta diferença nos resultados pode ser atribuída às variações observadas na qualidade das cimentações agrupadas no problema com 2 classes como “Isolamento Obtido”. Uma cimentação classificada como “Excelente” apresenta características bem distintas de uma cimentação classificada como “Média”. Apesar de, em ambas, os isolamentos requeridos terem sido obtidos, na cimentação “Excelente” mais de 70% do trecho cimentado apresenta no perfil acústico indícios de boa qualidade da cimentação, enquanto que na cimentação “Média” a proporção do intervalo com indícios de boa qualidade da cimentação é de até 30%. Dessa forma, entende-se que a cimentação classificada como “Média” não é um registro de cimentação que deve ser utilizado como referência de sucesso.

Esse trabalho mostrou que a utilização de modelos de predição pode direcionar a necessidade de perfilagem. No entanto, devido ao risco envolvido nesse tipo de operação e às drásticas consequências no caso de isolamentos não atendidos e não tratados de forma devida, é necessário o desenvolvimento de maiores estudos antes de decidir por não correr ferramentas de perfilagem para a avaliação da qualidade da cimentação.

Para continuidade dos estudos, pode-se realizar uma análise de risco de acordo com a probabilidade de classificação, fornecida por modelos como o de regressão logística. O resultado do modelo para cada registro fornecerá a probabilidade de sucesso da cimentação. Como medida de risco de se realizar ou não a perfilagem, a ideia é estabelecer um limite de corte para essa probabilidade.

Uma outra abordagem interessante seria a reclassificação dos registros, utilizando como resultado final da qualidade da cimentação a porcentagem de extensão com indício de isolamento verificada em perfil em relação à extensão cimentada. A análise da extensão bem cimentada deve ser realizada em conjunto com as informações de isolamento requeridos, ou seja, a identificação da qualidade da cimentação frente às formações de interesse para composição de conjuntos solidários de barreiras. Dessa maneira, poderiam ser utilizados modelos de regressão para predição. Nesse caso, poderia ser realizada também a análise de risco de acordo com a probabilidade determinada pelos modelos.

Após a conquista de um modelo de aprendizado com desempenho satisfatório para permitir a decisão pelo atestado de isolamento hidráulico, pode-se prosseguir com o desenvolvimento de um aplicativo que implante esse modelo em produção. Os dados da operação de cimentação, alimentados em um sistema capaz de fornecer a classificação esperada para aquela cimentação, ou no caso de se utilizar os modelos de probabilidade, a probabilidade de sucesso da operação e risco associado, subsidiariam as decisões dos especialistas. Os resultados coletados no ambiente de produção devem ser utilizados para refinamento contínuo do modelo

7 Referências bibliográficas

- [1] ACOSTA, J.; BARROSO, M.; MANDAL, B.; SOARES, D.; MILANKOVIC, A.; LIMA, L.; PIEDADE, T. New-Generation, Circumferential Ultrasonic Cement-Evaluation Tool for Thick Casings: Case Study in Ultradeepwater Well. Offshore Technology Conference. Outubro. 2017.
- [2] AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP. Resolução ANP N° 46. Acessado em março de 2017. Disponível em <http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2016/novembro&item=ranp-46--2016>
- [3] ALAWAD, M.N.J.; MOHAMMAD, H.A. Engineering Management and Inspection Schedule of Petroleum Well Integrity. JETEAS. 2016.
- [4] ALLOUCHE, M.; GUILOT, D.; HAYMAN, A.J.; BUTSCH, R.J.; MORRIS, C.W. Well Cementing, v.3.
- [5] AMERICAN PETROLEUM INSTITUTE. API SPEC 10B. Recommended Practice for Testing Well Cements. 2013
- [6] BELDJAZIA, A.; ALATOU, D. Precipitation Variability on the massif Forest of Mahouna (North Eastern-Algeria) from 1986 to 2010. International Journal of Management Sciences and Business Research. Vol 5, No 3. Março. 2016.
- [7] BP. Deepwater Horizon Accident Investigation Report. Set. 2010.
- [8] CASTILLO, E.; COBO, A.; GUTIÉRREZ, J.M.; PRUNEDA, E. Introduction to functional networks with applications. A neural based paradigm. 1999.
- [9] CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. Step-by-step data mining guide - CRISP-DM 1.0. SPSS. 2000.
- [10] DA FONSECA, T.C. Metodologia de Análise de Integridade para Projetos de Poços de Desenvolvimento da Produção. Dissertação de Mestrado em Ciências e Engenharia de Petróleo, Faculdade de Engenharia Mecânica e Instituto de Geociências, Universidade Estadual de Campinas, 2012.

- [11] DE LEPELEIRE, R. Perfis de Avaliação de Cimentação, ago. 2005. Petrobras.
- [12] EVSUKOFF, A.G. Ensinando Máquinas.
- [13] FRISCH, G.; GOODWIN, J. Cement Sheath Evaluation. 2003.
- [14] GAMA, J. Árvores de Decisão. Universidade do Porto – Faculdade de Ciências. Departamento de Ciência de Computadores. Acessado em Maio de 2018. Disponível em http://www.dcc.fc.up.pt/~ines/aulas/MIM/arvores_de_decisao.pdf
- [15] GERA, M. GOEL, S. Data Mining – Techniques, Methods and Algorithms: A Review on Tools and their Validity. International Journal of Computer Applications, Vol 113, No 18. Mar. 2015.
- [16] HAYKIN, S. Neural Networks – A Comprehensive Foundation. Edição 2. 1999.
- [17] IBRAHIM, D.S.; ALMAHDAWI, F.H.M. Addition of Super Absorbent Polymer for Upgrading of Cement Quality in Iraqy Oil Wells. *Iraqy Journal of Chemical and Petroleum Engineering*, v.17, n.3, p. 83-90, set. 2016.
- [18] IHS MARKIT. Petrodata Offshore Rig Day Rate Trends. Acessado em agosto de 2018. Disponível em <https://ihsmarkit.com/products/oil-gas-drilling-rigs-offshore-day-rates.html>
- [19] IZON, D.; DANENBERGER, E.P.; MAYES, M. Absence of fatalities in blowouts encouraging in MMS study of OCS incidents 1992-2006. Drilling Contractor. Jul. 2007.
- [20] KNIME Analytics Platform. Versão 2.11.3. KNIME AG.
- [21] LIMA, F.M. Avaliação do Comportamento Reológico de Pastas de Cimento para Poços de Petróleo com Adição de Plastificantes. Dissertação de Metrado em Ciência e Engenharia de Materiais, Universidade Federal do Rio Grande do Norte, 2006.
- [22] NIKAM, S.S. A Comparative Study of Classification Techniques in Data Mining Algorithms. Oriental Journal of Computer Science & Technology. Vol 8, No 1, Pags 13-19. Abr. 2015.
- [23] NORWEGIAN TECHNOLOGY STANDARDS INSTITUTION. NORSOK D-010. Well Integrity in drilling and well operations. Rev. 4. Jun. 2013.

- [24] PETROBRAS. Procelab – Manual de Procedimentos de Laboratório para Cimentação de Poços de Petróleo. 2013-2014.
- [25] PRATI, R.C.; BATISTA, G.E.A.P.A; MONARD, M.C. Curvas ROC para avaliação de classificadores. IEEE. 2008.
- [26] RIBEIRO, D.B.; CAMPOS, G.; PIEDADE, T.S.; COSTA, R.E.N. Cimentação de Poços de Petróleo. 2018. Petrobras.
- [27] SOUZA, G.S. Cimentação em Poços de Petróleo, v1. 2013. Petrobras.
- [28] VIGNES, B.; AADNOY, B.S. Well Integrity Issues Offshore Norway. SPE/IADC 112535. Mar. 2008.
- [29] ZUBEN, F.J.V; ATTUX, R.R.F. Árvores de Decisão. Unicamp. Disponível em: ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004_1s10/notas_de_aula/topico7_IA004_1s10.pdf. Acessado em agosto de 2018.

APÊNDICE A - Lista de atributos do BD1

VARIÁVEL	TIPO	DESCRIÇÃO
Poço	Nominal	Identificação do registro
Localização	Nominal	Região de construção do poço Domínio: <i>P1; P2; P3; P4; P5; P6; P7</i>
Fase	Nominal	Fase do poço a qual a cimentação avaliada está associada Domínio: <i>Intermediário; Liner Intermediário; Produção; Liner Produção</i>
Lâmina d'água	Inteiro	Extensão de lâmina d'água Domínio: <i>80 a 5030 (metros)</i>
Diâmetro do poço	Real	Diâmetro da broca que perfurou a fase do poço a qual a cimentação avaliada está associada. Está relacionada a espessura do anular que foi cimentado. Domínio: <i>8.5; 9; 12; 12.25; 14; 14.75; 16; 16.5; 17.5 (polegadas)</i>
Ângulo máximo da fase	Inteiro	Ângulo máximo de desvio da fase do poço a qual a cimentação avaliada está associada. Poços mais desviados tem maiores dificuldades de centralização da coluna de revestimento e conseqüentemente maior tendência a canalização da pasta de cimento. Domínio: <i>0 a 92 (graus)</i>

VARIÁVEL	TIPO	DESCRIÇÃO
Extensão centralizada	Inteiro	Extensão em que foram utilizados centralizadores na coluna de revestimento. A descentralização do revestimento propicia uma maior possibilidade de canalização da pasta de cimento. Domínio: 0 a 1520 (metros)
Número de centralizadores	Inteiro	Quantidade de centralizadores utilizados na coluna de revestimento. Uma pequena quantidade de centralizadores em grande extensão pode não ser suficiente para garantir a centralização da coluna de revestimento. Domínio: 0 a 90 (unidades)
Diâmetro dos centralizadores	Real	O diâmetro dos centralizadores utilizados deve estar de acordo com o diâmetro do poço no qual a coluna de revestimento deverá estar centralizada. Domínio: 8.5; 9 ; 11.75; 12.25; 12.5 ; 13.5 ; 13.75; 14.75; 16; 16.75; 17.5; 18
Tipo de fluido	Binária	O tipo de fluido no poço antes da cimentação deve ser quimicamente compatível com a pasta de cimento. Domínio: Aquoso; Não Aquoso
Fabricante do fluido	Nominal	Fabricante do fluido de perfuração, presente no poço antes da cimentação. Domínio: F1; F2; F3; F4; F5; F6
Densidade do fluido	Real	Densidade do fluido de perfuração, presente no poço antes da cimentação. Valores baixos de densidade do fluido facilitam a remoção deste pela pasta de cimento. Domínio: 8.6 a 14.5 (libras/galão)

VARIÁVEL	TIPO	DESCRIÇÃO
Viscosidade Plástica do fluido	Inteiro	Viscosidade plástica do fluido de perfuração, presente no poço antes da cimentação. Valores baixos de viscosidade plástica do fluido facilitam a remoção deste pela pasta de cimento. Domínio: 8 a 90 (<i>centipoise</i>)
Limite de Escoamento do fluido	Inteiro	Limite de escoamento do fluido de perfuração, presente no poço antes da cimentação. Valores baixos de limite de escoamento do fluido facilitam a remoção deste pela pasta de cimento. Domínio: 4 a 50 (<i>lbf/100ft²</i>)
Leitura de 0300 do fluido	Inteiro	Característica reológica do fluido de perfuração, presente no poço antes da cimentação. Medida de deflexão, em graus, decorrente de rotação do viscosímetro a 300 rpm. Domínio: 15 a 110 (<i>graus</i>)
Leitura de 03 do fluido	Inteiro	Característica reológica do fluido de perfuração, presente no poço antes da cimentação. Medida de deflexão, em graus, decorrente de rotação do viscosímetro a 3 rpm. Domínio: 2 a 20 (<i>graus</i>)
Temperatura de circulação do poço	Inteiro	A temperatura do poço influencia na taxa de desenvolvimento de resistência compressiva do cimento e no comportamento dos aditivos. Domínio: 65 a 225 (<i>graus Fahrenheit</i>)
Companhia de cimentação	Nominal	Companhia de serviço responsável pela cimentação avaliada. Domínio: A; B; C

VARIÁVEL	TIPO	DESCRIÇÃO
Tipo de cimento	Nominal	A Petrobras nas operações de cimentação dos revestimentos de poços petrolíferos utiliza 2 tipos de cimento: cimento básico desenvolvido para poços petrolíferos (Cimento Portland CPP - Classe G e Cimento Portland CPP - Classe Especial). Além desses, as companhias de cimentação disponibilizam <i>blends</i> (mistura de cimento com aditivos especiais) específicos para determinados cenários. Domínio: <i>G; Espuma; Blends proprietários companhias - 6</i>
Densidade da pasta	Real	Massa específica da pasta de cimento. É determinada de maneira a atender a janela operacional (pressões de poro e de fratura). Domínio: <i>11.5 a 16.5 (libras/galão)</i>
Filtrado	Inteiro	O filtrado de uma pasta de cimento é definido como a perda de fluido durante 30 minutos, em condições pré-estabelecidas. Deseja-se que a quantidade de água perdida para formação seja o menor possível. Domínio: <i>0 a 590 (cm³/30 min)</i>
Água livre	Inteiro	Determina o volume de água acumulada acima da pasta de cimento após duas horas de repouso. Domínio: <i>0 a 3.8 (%)</i>
Consistometria (100 UC)	Inteiro	Tempo, em minutos, em que a pasta de cimento permanece fluida em determinadas condições de pressão e temperatura. Domínio: <i>100 a 285 (minutos)</i>

VARIÁVEL	TIPO	DESCRIÇÃO
Resistência compressiva 24h	Inteiro	Resistência à compressão, em psi, que a pasta de cimento atinge após 24 horas, sob determinadas condições de pressão e temperatura Domínio: 600 a 6500 (<i>minutos</i>)
Leitura de 0300 da pasta	Inteiro	Característica reológica da pasta de cimento. Medida de deflexão, em graus, decorrente de rotação do viscosímetro a 300 rpm. Domínio: 30 a 300 (<i>graus</i>)
Leitura de 03 da pasta	Inteiro	Característica reológica da pasta de cimento. Medida de deflexão, em graus, decorrente de rotação do viscosímetro a 3 rpm. Domínio: 3 a 65 (<i>graus</i>)
Gel inicial	Inteiro	Resistência da pasta de cimento ao reiniciar um movimento depois de parada de 10 segundos. Domínio: 3 a 40 (<i>graus</i>)
Gel final	Inteiro	Resistência da pasta de cimento ao reiniciar um movimento depois de parada de 10 minutos. Domínio: 5 a 90 (<i>graus</i>)
Viscosidade plástica da pasta	Inteiro	Viscosidade plástica da pasta de cimento. Deve ser baixa para facilitar sua penetração nos anulares, oferecer boas condições de bombeabilidade e aderir fortemente à formação e ao revestimento. Domínio: 8 a 450 (<i>centipoise</i>)

VARIÁVEL	TIPO	DESCRIÇÃO
Limite de escoamento da pasta	Inteiro	Limite de escoamento da pasta de cimento. Deve ser superior ao limite de escoamento do fluido que ela deslocará, obedecendo a hierarquia reológica. Domínio: 4 a 285 (lbf/100ft ²)
Presença de sílica	Binário	Presença de sílica, aditivo para prevenir a retrogressão da pasta de cimento.
Presença de controlador de filtrado	Binário	Presença de aditivo controlador de filtrado na pasta de cimento.
Presença de bloqueador de gás	Binário	Presença de aditivo controlador de migração de gás na pasta de cimento.
Presença de retardador	Binário	Presença de aditivo retardador de desenvolvimento de resistência compressiva na pasta de cimento.
Presença de agente expansivo	Binário	Presença de aditivo responsável pela expansão da pasta de cimento.
Presença de dispersante	Binário	Presença de aditivo dispersante na pasta de cimento.
Presença de viscosificante	Binário	Presença de aditivo viscosificante na pasta de cimento.
Presença de controlador de água livre	Binário	Presença de aditivo controlador de água livre na pasta de cimento.

VARIÁVEL	TIPO	DESCRIÇÃO
Extensão cimentada	Inteiro	A extensão cimentada deve ser projetada a fim de serem cobertos todos os isolamentos requeridos para a fase. Domínio: <i>160 a 1530 (metros)</i>
Volume de colchão lavador	Inteiro	Colchões lavadores agem quimicamente para afinar e dispersar o fluido presente no anular antes da cimentação, facilitando assim sua remoção. Domínio: <i>0 a 460 (barris)</i>
Volume de colchão espaçador	Inteiro	Colchões espaçadores, além de agir quimicamente para dispersar o fluido presente no anular antes da cimentação, proporcionam forças adicionais de flutuação e de arraste que facilitam a remoção do fluido. Domínio: <i>0 a 280 (barris)</i>
Volume de pasta	Inteiro	Volume de pasta bombeado para o poço, a fim de atingir o topo de cimento projetado. Domínio: <i>60 a 500 (barris)</i>
Volume de excesso	Inteiro	Volume de pasta bombeado em excesso, ou seja, além do volume necessário para atingir o topo de cimento projetado. O Excesso de pasta permite a compensação de arrombamentos e perdas durante a cimentação. Domínio: <i>0 a 200 (barris)</i>
Vazão de deslocamento	Inteiro	Testes de campo e simuladores indicam que melhores eficiências de remoção do fluido presente no anular antes da cimentação são alcançadas quando a pasta de cimento é deslocada a altas vazões. Domínio: <i>7 a 18 (barris/minuto)</i>

VARIÁVEL	TIPO	DESCRIÇÃO
Separador mecânico na frente da pasta	Binário	O separador mecânico entre colchão e pasta de cimento impede a contaminação da pasta de cimento garantindo a manutenção das propriedades requeridas para a pasta, aumentando a possibilidade de atingir os objetivos propostos.
Volume de perda	Inteiro	Volume de fluido perdido para a formação durante a operação de cimentação. Quanto maior o volume perdido, mais difícil atingir os objetivos propostos para a cimentação. <i>Domínio: 0 a 1880 (barris)</i>
Qualidade da cimentação	Nominal	VARIÁVEL ALVO. Qualidade da cimentação obtida por meio de perfis acústicos.

APÊNDICE B - Lista de atributos do BD2

VARIÁVEL	TIPO	DESCRIÇÃO
Poço	Nominal	Identificação do registro.
Poço calibrado	Binária	O perfil caliper é importante para dimensionamento do volume de pasta a ser bombeado para o poço e projeto de centralização e indicação de possíveis arrombamentos.
Pressões de bombeio conforme simulado	Binária	Pressões acima do simulado podem indicar "embuchamento" (alteração na geometria do poço) ou fluido mal condicionado, que podem influenciar na colocação da pasta.
Arraste com diferenças até 20% do simulado	Binária	Verificar arrastes acima do simulado que possam indicar restrição no poço, que podem danificar ou retirar os centralizadores da posição instalada, principalmente em revestimentos sem luvas.
Ausência de perdas durante a descida do revestimento e circulação para cimentar	Binária	As perdas durante a descida ou circulações podem indicar que a extensão de pasta pode ser comprometida e o volume do colchão perdido desfavoreça a limpeza. Além disso o regime de escoamento pode prejudicar a limpeza do poço.
Ausência de ganho durante a descida do revestimento e circulação para cimentar	Binária	Ganhos existentes durante as circulações podem se repetir durante o aguardo de pega e consequentemente gerar contaminação da pasta de cimento.

VARIÁVEL	TIPO	DESCRIÇÃO
Liberação do plugue de topo com o volume esperado	Binária	A liberação do plugue de topo indica a separação mecânica entre a pasta de cimento e o fluido de deslocamento, preservando a qualidade da pasta de cimento no shoe-track.
Batida de plug de topo com o volume esperado	Binária	A batida do plugue de topo indica o posicionamento de todo o volume de pasta planejado no anular.
Dificuldades no ajuste do peso da pasta	Binária	A densidade da pasta de cimento misturada no campo deve ser o mais próximo possível do desenho de pasta realizado em laboratório e considerado nas simulações hidráulica, térmica e de remoção, de forma a validar todas as propriedades físicas/químicas e atendimento aos objetivos da operação.
Vazão de deslocamento conforme planejado	Binária	O schedule de vazões realizado deve ser o mais próximo possível do considerado nas simulações de hidráulica, térmica e de remoção, de forma a validar atendimento aos objetivos da operação.
Teste do revestimento positivo contra o plugue de topo	Binária	O teste de pressão após a batida do plugue comprova ausência de vazamento pelo shoe-track.
Diferencial pressão compatível com topo cimento esperado	Binária	O crescimento de pressão no final de deslocamento indica a elevação pelo anular dos colchões pesados e pasta de cimento no anular.

VARIÁVEL	TIPO	DESCRIÇÃO
Ausência de flow back ao final da operação	Binária	A ausência de flow-back ao final da cimentação indica que os dispositivos flutuantes atuaram corretamente para impedir o retorno do cimento para o interior do revestimento, garantindo o posicionamento no anular.
Ausência de Perdas observadas durante cimentação e deslocamento	Binária	A ausência de perdas durante a cimentação indica que os volumes de colchões e pasta bombeados foram posicionados no anular e os regimes de fluxos são o mais próximo possível dos considerados nas simulações de hidráulica, térmica e de remoção.
Ausência de Ganho observado durante cimentação e deslocamento	Binária	A ausência de ganho observado durante a cimentação indica que não há entrada de fluidos no poço (oriundos da formação ou devolução de fluido perdido para a formação) que possa comprometer a obtenção de isolamentos.
Qualidade da cimentação	Nominal	VARIÁVEL ALVO. Qualidade da cimentação obtida por meio de perfis acústicos.

APÊNDICE C - Lista de atributos do BD3

VARIÁVEL	TIPO	DESCRIÇÃO
Poço	Nominal	Identificação do registro.
Extensão centralizada	Inteiro	Extensão em que foram utilizados centralizadores na coluna de revestimento. A descentralização do revestimento propicia uma maior possibilidade de canalização da pasta de cimento. Domínio: 0 a 1520 (metros)
Número de centralizadores	Inteiro	Quantidade de centralizadores utilizados na coluna de revestimento. Uma pequena quantidade de centralizadores em grande extensão pode não ser suficiente para garantir a centralização da coluna de revestimento. Domínio: 0 a 90 (unidades)
Densidade do fluido	Real	Valores baixos de densidade do fluido no poço facilitam a remoção deste pela pasta de cimento. Domínio: 8.6 a 14.5 (libras/galão)
Viscosidade Plástica do fluido	Inteiro	Valores baixos de viscosidade plástica do fluido no poço facilitam a remoção deste pela pasta de cimento. Domínio: 8 a 90 (centipoise)

VARIÁVEL	TIPO	DESCRIÇÃO
Limite de Escoamento do fluido	Inteiro	Valores baixos de limite de escoamento do fluido no poço facilitam a remoção deste pela pasta de cimento. Domínio: <i>4 a 50 (lbf/100ft²)</i>
Leitura de 0300 do fluido	Inteiro	Característica reológica do fluido de perfuração, presente no poço antes da cimentação. Medida de deflexão, em graus, decorrente de rotação do viscosímetro a 300 rpm. Domínio: <i>15 a 110 (graus)</i>
Leitura de 03 do fluido	Inteiro	Característica reológica do fluido de perfuração, presente no poço antes da cimentação. Medida de deflexão, em graus, decorrente de rotação do viscosímetro a 3 rpm. Domínio: <i>2 a 20 (graus)</i>
Extensão cimentada	Inteiro	A extensão cimentada deve ser projetada a fim de serem cobertos todos os isolamentos requeridos para a fase. Domínio: <i>160 a 1530 (metros)</i>
Volume de colchão lavador	Inteiro	Colchões lavadores agem quimicamente para afinar e dispersar o fluido presente no anular antes da cimentação, facilitando assim sua remoção. Domínio: <i>0 a 460 (barris)</i>

VARIÁVEL	TIPO	DESCRIÇÃO
Volume de colchão espaçador	Inteiro	Colchões espaçadores, além de agir quimicamente para dispersar o fluido presente no anular antes da cimentação, proporcionam forças adicionais de flutuação e de arraste que facilitam a remoção do fluido. Domínio: 0 a 280 (barris)
Volume de pasta	Inteiro	Volume de pasta bombeado para o poço, a fim de atingir o topo de cimento projetado. Domínio: 60 a 500 (barris)
Volume de excesso	Inteiro	Volume de pasta bombeado em excesso, ou seja, além do volume necessário para atingir o topo de cimento projetado. O Excesso de pasta permite a compensação de arrombamentos e perdas durante a cimentação. Domínio: 0 a 200 (barris)
Vazão de deslocamento	Inteiro	Testes de campo e simuladores indicam que melhores eficiências de remoção do fluido presente no anular antes da cimentação são alcançadas quando a pasta de cimento é deslocada a altas vazões. Domínio: 7 a 18 (barris/minuto)

VARIÁVEL	TIPO	DESCRIÇÃO
Separador mecânico na frente da pasta	Binário	O separador mecânico entre colchão e pasta de cimento impede a contaminação da pasta de cimento garantindo a manutenção das propriedades requeridas para a pasta, aumentando a possibilidade de atingir os objetivos propostos.
Volume de perda	Inteiro	Volume de fluido perdido para a formação durante a operação de cimentação. Quanto maior o volume perdido, mais difícil atingir os objetivos propostos para a cimentação. Domínio: 0 a 1880 (barris)
Poço calibrado	Binária	O perfil caliper é importante para dimensionamento do volume de pasta a ser bombeado para o poço e projeto de centralização e indicação de possíveis arrombamentos.
Pressões de bombeio conforme simulado	Binária	Pressões acima do simulado podem indicar "embuchamento" (alteração na geometria do poço) ou fluido mal condicionado, que podem influenciar na colocação da pasta.
Arraste com diferenças até 20% do simulado	Binária	Verificar arrastes acima do simulado que possam indicar restrição no poço, que podem danificar ou retirar os centralizadores da posição instalada, principalmente em revestimentos sem luvas.

VARIÁVEL	TIPO	DESCRIÇÃO
Ausência de perdas durante a descida do revestimento e circulação para cimentar	Binária	As perdas durante a descida ou circulações podem indicar que a extensão de pasta pode ser comprometida e o volume do colchão perdido desfavoreça a limpeza. Além disso o regime de escoamento pode prejudicar a limpeza do poço.
Ausência de ganho durante a descida do revestimento e circulação para cimentar	Binária	Ganhos existentes durante as circulações podem se repetir durante o aguardo de pega e consequentemente gerar contaminação da pasta de cimento.
Liberação do plugue de topo com o volume esperado	Binária	A liberação do plugue de topo indica a separação mecânica entre a pasta de cimento e o fluido de deslocamento, preservando a qualidade da pasta de cimento no shoe-track.
Batida de plug de topo com o volume esperado	Binária	A batida do plugue de topo indica o posicionamento de todo o volume de pasta planejado no anular.
Dificuldades no ajuste do peso da pasta	Binária	A densidade da pasta de cimento misturada no campo deve ser o mais próximo possível do desenho de pasta realizado em laboratório e considerado nas simulações hidráulica, térmica e de remoção, de forma a validar todas as propriedades físicas/químicas e atendimento aos objetivos da operação

VARIÁVEL	TIPO	DESCRIÇÃO
Vazão de deslocamento conforme planejado	Binária	O schedule de vazões realizado deve ser o mais próximo possível do considerado nas simulações de hidráulica, térmica e de remoção, de forma a validar atendimento aos objetivos da operação.
Teste do revestimento positivo contra o plugue de topo	Binária	O teste de pressão após a batida do plugue comprova ausência de vazamento pelo shoe-track.
Diferencial pressão compatível com topocimento esperado	Binária	O crescimento de pressão no final de deslocamento indica a elevação pelo anular dos colchões pesados e pasta de cimento no anular.
Ausência de flow back ao final da operação	Binária	A ausência de flow-back ao final da cimentação indica que os dispositivos flutuantes atuaram corretamente para impedir o retorno do cimento para o interior do revestimento, garantindo o posicionamento no anular.
Ausência de Perdas observadas durante cimentação e deslocamento	Binária	A ausência de perdas durante a cimentação indica que os volumes de colchões e pasta bombeados foram posicionados no anular e os regimes de fluxos são o mais próximo possível dos considerados nas simulações de hidráulica, térmica e de remoção.

VARIÁVEL	TIPO	DESCRIÇÃO
Ausência de Ganho observado durante cimentação e deslocamento	Binária	A ausência de ganho observado durante a cimentação indica que não há entrada de fluidos no poço (oriundos da formação ou devolução de fluido perdido para a formação) que possa comprometer a obtenção de isolamentos.
Qualidade da cimentação	Nominal	VARIÁVEL ALVO. Qualidade da cimentação obtida por meio de perfis acústicos.

APÊNDICE D – Resumo das informações presente em BD1, BD2 e BD3

Nº	Variáveis	BD1	BD2	BD3
1	Identificação do registro	x	x	x
2	Localização do poço	x		
3	Fase do poço	x		
4	Extensão da lâmina d'água	x		
5	Diâmetro do poço	x		
6	Temperatura de circulação do poço	x		
7	Inclinação máximo da fase	x		
8	Centralização do revestimento	x	x	x
9	Tipo de fluido no poço	x		
10	Fabricante do fluido no poço	x		
11	Densidade do fluido no poço	x		x
12	VP do fluido no poço	x		x
13	LE do fluido no poço	x		x
14	L300 do fluido no poço	x		x
15	L3 do fluido no poço	x		x
16	Companhia de Cimentação	x		
17	Tipo de Cimento	x		
18	Densidade da pasta	x		
19	Filtrado	x		
20	Água Livre	x		
21	Consistometria 100 UC	x		
22	Resistência compressiva 24h	x		
23	L300 da pasta	x		
24	L3 da pasta	x		
25	Gi	x		
26	Gf	x		
27	VP da pasta	x		
28	LE da pasta	x		
29	Presença de Sílica	x		
30	Presença de Controlador de Filtrado	x		
31	Presença de Bloqueador de Gás	x		
32	Presença de Retardador	x		
33	Presença de Agente Expansivo	x		
34	Presença de Dispersante	x		
35	Presença de Viscosificante	x		
36	Presença de Controlador de água livre	x		
37	Extensão Cimentada	x		x

Nº	Variáveis	BD1	BD2	BD3
38	Volume de Colchão Lavador	x		x
39	Volume de Colchão Espaçador	x		x
40	Volume de pasta	x		x
41	Volume de Excesso	x		x
42	Separador mecânico na frente da pasta	x		x
43	Análise de perdas de circulação	x	x	x
44	Análise da calibração do poço		x	x
45	Análise das pressões de bombeio		x	x
46	Análise de arraste do revestimento		x	x
47	Análise de ganho de circulação		x	x
48	Análise da liberação do plugue de topo		x	x
49	Análise da Batida de Plug de topo		x	x
50	Ajuste do peso da pasta		x	x
51	Análise da vazão de deslocamento	x	x	x
52	Resultado do teste de pressão no revestimento		x	x
53	Análise do diferencial de pressão		x	x
54	Análise de flow back ao final da operação		x	x
55	Qualidade da cimentação	x	x	x

APÊNDICE E – Poços com Falha de Isolamento

A tabela neste apêndice apresenta a lista de falhas na operação de cimentação dos poços em que foram constatadas falhas de isolamento hidráulico por análise de perfil acústico.

Poço1	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço2	Perdas observadas durante as circulações para cimentar
	Falha na liberação do plugue de topo
	Falha na batida do plugue de topo
	Dificuldades no ajuste do peso da pasta
	Vazão de deslocamento abaixo do planejado
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço3	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço4	Diferencial pressão incompatível com o topo de cimento esperado
	Ganho observado durante cimentação e deslocamento
Poço5	Arrastes com diferenças maiores que 20% do simulado
	Falha na liberação do plugue de topo
	Falha na batida do plugue de topo
	Perdas observadas durante cimentação e deslocamento
Poço6	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço7	Perdas observadas durante as circulações para cimentar
	Falha na liberação do plugue de topo
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço8	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço9	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço10	Perdas observadas durante as circulações para cimentar
Poço12	Perdas observadas durante as circulações para cimentar
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço22	Perdas observadas durante as circulações para cimentar
	Dificuldades no ajuste do peso da pasta
Poço32	Perdas observadas durante as circulações para cimentar
	Falha na liberação do plugue de topo

	Falha na batida do plugue de topo
	Presença de flow back ao final da operação
	Perdas observadas durante cimentação e deslocamento
	Ganho observado durante cimentação e deslocamento
Poço49	Falha na liberação do plugue de topo
	Falha na batida do plugue de topo
	Diferencial pressão incompatível com o topo de cimento esperado
Poço52	Vazão de deslocamento abaixo do planejado
	Presença de flow back ao final da operação
	Perdas observadas durante cimentação e deslocamento
Poço76	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço101	Arrastes com diferenças maiores que 20% do simulado
	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço105	Perdas observadas durante cimentação e deslocamento
Poço126	Pressões de bombeio com diferenças maiores que 20% em relação simulado
	Arrastes com diferenças maiores que 20% do simulado
	Falha na batida do plugue de topo
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço131	Perdas observadas durante as circulações para cimentar
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço146	Arrastes com diferenças maiores que 20% do simulado
	Perdas observadas durante cimentação e deslocamento
Poço152	Perdas observadas durante as circulações para cimentar
	Ganho observado durante as circulações para cimentar
	Falha na batida do plugue de topo
	Dificuldades no ajuste do peso da pasta
	Perdas observadas durante cimentação e deslocamento
Poço163	Perdas observadas durante as circulações para cimentar
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço201	Poço não calibrado
Poço204	Perdas observadas durante as circulações para cimentar
	Falha na batida do plugue de topo
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço215	Poço não calibrado
Poço221	Perdas observadas durante as circulações para cimentar
	Diferencial pressão incompatível com o topo de cimento esperado

	Perdas observadas durante cimentação e deslocamento
Poço236	Falha na liberação do plugue de topo
	Falha na batida do plugue de topo
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento
Poço237	Perdas observadas durante as circulações para cimentar
	Perdas observadas durante cimentação e deslocamento
Poço245	Perdas observadas durante as circulações para cimentar
	Falha na liberação do plugue de topo
	Falha na batida do plugue de topo
	Diferencial pressão incompatível com o topo de cimento esperado
	Perdas observadas durante cimentação e deslocamento

APÊNDICE F – Matrizes de confusão

1- Classificação do Banco de Dados 1

a. Naive Bayes

Tabela 22 - Matriz de confusão para Naive Bayes em BD1 com 2 classes

	Isolamento (predita)	Falha Isolamento (predita)
Isolamento (real)	397	38
Falha Isolamento (real)	20	6

Tabela 23 - Matriz de Confusão para Naive Bayes em BD1 com 4 classes

	Excelente (predita)	Boa (predita)	Média (predita)	Ruim (predita)
Excelente (real)	65	96	26	21
Boa (real)	44	64	22	10
Média (real)	17	47	18	5
Ruim (real)	10	9	1	6

b. Árvore de Decisão

Tabela 24 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, sem poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	414	16
<i>Falha Isolamento (real)</i>	23	1

Tabela 25 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, sem poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	416	15
<i>Falha Isolamento (real)</i>	23	2

Tabela 26 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, com poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	431	3
<i>Falha Isolamento (real)</i>	26	0

Tabela 27 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gini Index, com poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	430	5
<i>Falha Isolamento (real)</i>	26	0

Tabela 28 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, sem poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	425	10
<i>Falha Isolamento (real)</i>	25	1

Tabela 29 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, sem poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	423	12
<i>Falha Isolamento (real)</i>	25	1

Tabela 30 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, com poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	432	3
<i>Falha Isolamento (real)</i>	24	2

Tabela 31 - Matriz de Confusão para Árvores de Decisão em BD1 com 2 classes – Gain Ratio, com poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	433	2
<i>Falha Isolamento (real)</i>	25	1

Tabela 32 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes – Gini Index, sem poda, ponto médio de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	97	76	28	5
<i>Boa (real)</i>	66	46	20	5
<i>Média (real)</i>	34	24	25	2
<i>Ruim (real)</i>	15	8	0	2

*Tabela 33 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes – Gini
Index, sem poda, máximo ganho de particionamento*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	108	65	28	4
<i>Boa (real)</i>	68	45	18	7
<i>Média (real)</i>	31	28	22	5
<i>Ruim (real)</i>	8	9	4	4

*Tabela 34 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes – Gini
Index, com poda, ponto médio de particionamento*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	126	49	10	1
<i>Boa (real)</i>	96	26	15	2
<i>Média (real)</i>	54	16	15	1
<i>Ruim (real)</i>	19	6	0	1

Tabela 35 - Matriz de Confusão para Árvores de Decisão em BDI com 4 classes – Gini Index, com poda, máximo ganho de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	164	28	14	0
<i>Boa (real)</i>	103	24	10	2
<i>Média (real)</i>	50	20	14	2
<i>Ruim (real)</i>	17	8	1	0

Tabela 36 - Matriz de Confusão para Árvores de Decisão em BDI com 4 classes – Gain Ratio, sem poda, ponto médio de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	195	7	2	4
<i>Boa (real)</i>	126	7	7	0
<i>Média (real)</i>	77	4	5	1
<i>Ruim (real)</i>	26	0	0	0

Tabela 37 - Matriz de Confusão para Árvores de Decisão em BDI com 4 classes – Gain

Ratio, sem poda, máximo ganho de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	195	9	2	2
<i>Boa (real)</i>	123	11	5	1
<i>Média (real)</i>	72	9	4	2
<i>Ruim (real)</i>	25	0	1	0

Tabela 38 - Matriz de Confusão para Árvores de Decisão em BDI com 4 classes – Gain

Ratio, com poda, ponto médio de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	204	3	1	0
<i>Boa (real)</i>	132	7	1	0
<i>Média (real)</i>	84	3	0	0
<i>Ruim (real)</i>	26	0	0	0

Tabela 39 - Matriz de Confusão para Árvores de Decisão em BD1 com 4 classes – Gain
Ratio com poda, máximo ganho de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	203	4	1	0
<i>Boa (real)</i>	133	7	0	0
<i>Média (real)</i>	84	2	1	0
<i>Ruim (real)</i>	26	0	0	0

c. Tree Ensemble

Tabela 40 - Matriz de Confusão para Tree Ensemble em BD1 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	406	29
<i>Falha Isolamento (real)</i>	22	4

Tabela 41 - Matriz de Confusão para Tree Ensemble em BD1 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	98	66	30	14
<i>Boa (real)</i>	62	41	23	14
<i>Média (real)</i>	31	27	23	6
<i>Ruim (real)</i>	9	10	4	3

d. Redes Neurais

Tabela 42 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD1 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	420	15
<i>Falha Isolamento (real)</i>	23	3

Tabela 43 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD1 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	419	16
<i>Falha Isolamento (real)</i>	24	2

Tabela 44 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD1 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	422	13
<i>Falha Isolamento (real)</i>	23	3

Tabela 45 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD1 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	419	16
<i>Falha Isolamento (real)</i>	23	3

Tabela 46 - Matriz de Confusão para o Redes Neurais (1 camada com 10 neurônios) em BD1 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	146	47	10	5
<i>Boa (real)</i>	93	33	11	3
<i>Média (real)</i>	53	18	13	3
<i>Ruim (real)</i>	15	6	3	2

Tabela 47 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD1 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	137	55	12	4
<i>Boa (real)</i>	86	42	9	3
<i>Média (real)</i>	51	19	12	5
<i>Ruim (real)</i>	16	4	3	3

Tabela 48 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD1 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	140	49	13	6
<i>Boa (real)</i>	76	42	17	5
<i>Média (real)</i>	52	16	15	4
<i>Ruim (real)</i>	13	4	4	5

Tabela 49 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD1 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	135	46	23	4
<i>Boa (real)</i>	88	38	12	2
<i>Média (real)</i>	52	17	16	2
<i>Ruim (real)</i>	15	4	4	3

2- Classificação do Banco de Dados 2

a. Naive Bayes

Tabela 50 - Matriz de confusão para Naive Bayes em BD2 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	163	54
<i>Falha Isolamento (real)</i>	13	17

Tabela 51 - Matriz de Confusão para Naive Bayes em BD2 com 4 classes

	Excelente (predita)	Boa (predita)	Média (predita)	Ruim (predita)
Excelente (real)	91	14	7	29
Boa (real)	21	6	4	23
Média (real)	8	2	2	10
Ruim (real)	3	2	4	21

b. Árvore de Decisão

Tabela 52 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gini Index, sem poda

	Isolamento (predita)	Falha Isolamento (predita)
Isolamento (real)	209	8
Falha Isolamento (real)	22	8

Tabela 53 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gini Index, com poda

	Isolamento (predita)	Falha Isolamento (predita)
Isolamento (real)	210	7
Falha Isolamento (real)	22	8

Tabela 54 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gain Ratio, sem poda

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	208	9
<i>Falha Isolamento (real)</i>	22	8

Tabela 55 - Matriz de Confusão para Árvores de Decisão em BD2 com 2 classes – Gain Ratio, com poda

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	217	0
<i>Falha Isolamento (real)</i>	30	0

Tabela 56 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes– Gini Index, sem poda

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	128	7	1	5
<i>Boa (real)</i>	36	5	0	3
<i>Média (real)</i>	14	3	0	5
<i>Ruim (real)</i>	12	5	0	13

*Tabela 57 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes – Gini
Index, com poda*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	135	1	0	5
<i>Boa (real)</i>	53	0	0	1
<i>Média (real)</i>	18	1	0	3
<i>Ruim (real)</i>	19	0	0	11

*Tabela 58 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes – Gain
Ratio, sem poda*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	129	5	1	6
<i>Boa (real)</i>	15	2	0	5
<i>Média (real)</i>	44	4	1	5
<i>Ruim (real)</i>	13	2	2	13

Tabela 59 - Matriz de Confusão para Árvores de Decisão em BD2 com 4 classes – Gain Ratio, com poda

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	139	0	0	2
<i>Boa (real)</i>	53	0	0	1
<i>Média (real)</i>	21	0	0	1
<i>Ruim (real)</i>	28	0	0	2

c. Tree Ensemble

Tabela 60 - Matriz de Confusão para Tree Ensemble em BD2 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	204	13
<i>Falha Isolamento (real)</i>	19	11

Tabela 61 - Matriz de Confusão para Tree Ensemble em BD2 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	128	5	3	5
<i>Boa (real)</i>	42	3	4	5
<i>Média (real)</i>	15	2	3	2
<i>Ruim (real)</i>	13	4	3	10

d. Redes Neurais

Tabela 62 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD2 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	207	10
<i>Falha Isolamento (real)</i>	19	11

Tabela 63 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD2 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	207	10
<i>Falha Isolamento (real)</i>	21	9

Tabela 64 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD2 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	208	9
<i>Falha Isolamento (real)</i>	21	9

Tabela 65 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD2 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	208	9
<i>Falha Isolamento (real)</i>	20	10

Tabela 66 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD2 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	125	12	0	4
<i>Boa (real)</i>	47	6	0	1
<i>Média (real)</i>	15	3	0	4
<i>Ruim (real)</i>	14	4	1	11

Tabela 67 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD2 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	119	17	1	4
<i>Boa (real)</i>	45	6	0	3
<i>Média (real)</i>	13	5	0	4
<i>Ruim (real)</i>	11	7	3	9

Tabela 68 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD2 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	121	14	2	4
<i>Boa (real)</i>	46	6	0	2
<i>Média (real)</i>	15	3	0	4
<i>Ruim (real)</i>	10	9	2	9

Tabela 69 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD2 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	130	6	2	3
<i>Boa (real)</i>	44	8	0	2
<i>Média (real)</i>	16	2	2	2
<i>Ruim (real)</i>	11	4	2	13

3- Classificação do Banco de Dados 3

a. Naive Bayes

Tabela 70 - Matriz de confusão para Naive Bayes em BD3 com 2 classes

	Isolamento (predita)	Falha Isolamento (predita)
Isolamento (real)	172	40
Falha Isolamento (real)	9	19

Tabela 71 - Matriz de Confusão para Naive Bayes em BD3 com 4 classes

	Excelente (predita)	Boa (predita)	Média (predita)	Ruim (predita)
Excelente (real)	71	38	6	22
Boa (real)	21	17	7	9
Média (real)	8	5	2	6
Ruim (real)	2	5	2	19

b. Árvore de Decisão

Tabela 72 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes– Gini Index, sem poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	193	19
<i>Falha Isolamento (real)</i>	18	10

Tabela 73 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gini Index, sem poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	193	19
<i>Falha Isolamento (real)</i>	16	12

Tabela 74 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gini Index, com poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	199	13
<i>Falha Isolamento (real)</i>	22	6

Tabela 75 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gini Index, com poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	200	12
<i>Falha Isolamento (real)</i>	20	8

Tabela 76 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gain Ratio, sem poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	190	22
<i>Falha Isolamento (real)</i>	19	9

Tabela 77 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gain Ratio, sem poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	194	18
<i>Falha Isolamento (real)</i>	21	7

Tabela 78 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gain Ratio, com poda, ponto médio de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	212	0
<i>Falha Isolamento (real)</i>	28	0

Tabela 79 - Matriz de Confusão para Árvores de Decisão em BD3 com 2 classes – Gain Ratio, com poda, máximo ganho de particionamento

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	212	0
<i>Falha Isolamento (real)</i>	28	0

Tabela 80 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gini Index, sem poda, ponto médio de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	87	32	10	8
<i>Boa (real)</i>	29	16	4	5
<i>Média (real)</i>	11	5	1	4
<i>Ruim (real)</i>	9	8	3	8

*Tabela 81 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gini
Index, sem poda, máximo ganho de particionamento*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	82	37	11	7
<i>Boa (real)</i>	34	10	3	7
<i>Média (real)</i>	12	3	1	5
<i>Ruim (real)</i>	12	4	4	8

*Tabela 82 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gini
Index, com poda, ponto médio de particionamento*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	120	5	0	12
<i>Boa (real)</i>	44	2	0	8
<i>Média (real)</i>	16	0	0	5
<i>Ruim (real)</i>	17	2	0	9

Tabela 83 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gini Index, com poda, máximo ganho de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	125	2	0	10
<i>Boa (real)</i>	49	0	0	5
<i>Média (real)</i>	16	1	0	4
<i>Ruim (real)</i>	20	1	0	7

Tabela 84 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gain Ratio, sem poda, ponto médio de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	73	42	10	12
<i>Boa (real)</i>	35	11	3	5
<i>Média (real)</i>	14	2	0	5
<i>Ruim (real)</i>	8	9	1	10

Tabela 85 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gain

Ratio, sem poda, máximo ganho de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	90	32	3	12
<i>Boa (real)</i>	32	8	3	11
<i>Média (real)</i>	11	3	2	5
<i>Ruim (real)</i>	8	9	3	9

Tabela 86 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gain

Ratio, com poda, ponto médio de particionamento

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	136	0	1	0
<i>Boa (real)</i>	54	0	0	0
<i>Média (real)</i>	21	0	0	0
<i>Ruim (real)</i>	27	0	1	0

*Tabela 87 - Matriz de Confusão para Árvores de Decisão em BD3 com 4 classes – Gain
Ratio com poda, máximo ganho de particionamento*

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	136	1	0	0
<i>Boa (real)</i>	54	0	0	0
<i>Média (real)</i>	21	0	0	0
<i>Ruim (real)</i>	28	0	0	0

c. Tree Ensemble

Tabela 88 - Matriz de Confusão para Tree Ensemble em BD3 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	194	18
<i>Falha Isolamento (real)</i>	22	6

Tabela 89 - Matriz de Confusão para Tree Ensemble em BD3 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	83	41	3	3
<i>Boa (real)</i>	31	13	6	4
<i>Média (real)</i>	9	6	3	3
<i>Ruim (real)</i>	4	7	6	11

d. **Redes Neurais**

Tabela 90 - Matriz de Confusão para Redes Neurais (1 camada com 10 neurônios) em BD3 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	201	11
<i>Falha Isolamento (real)</i>	19	9

Tabela 91 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD3 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	200	12
<i>Falha Isolamento (real)</i>	17	11

Tabela 92 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD3 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	196	16
<i>Falha Isolamento (real)</i>	17	11

Tabela 93 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD3 com 2 classes

	<i>Isolamento (predita)</i>	<i>Falha Isolamento (predita)</i>
<i>Isolamento (real)</i>	201	11
<i>Falha Isolamento (real)</i>	17	11

Tabela 94 - Matriz de Confusão para o modelo de Redes Neurais (1 camada com 10 neurônios) em BD3 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	106	25	3	3
<i>Boa (real)</i>	35	12	3	4
<i>Média (real)</i>	12	4	1	4
<i>Ruim (real)</i>	5	7	3	13

Tabela 95 - Matriz de Confusão para Redes Neurais (1 camada com 20 neurônios) em BD3 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	107	20	4	6
<i>Boa (real)</i>	37	15	1	1
<i>Média (real)</i>	13	2	3	3
<i>Ruim (real)</i>	10	2	3	13

Tabela 96 - Matriz de Confusão para Redes Neurais (2 camadas com 10 neurônios cada) em BD3 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	99	27	6	5
<i>Boa (real)</i>	35	13	2	4
<i>Média (real)</i>	14	2	2	3
<i>Ruim (real)</i>	9	5	3	11

Tabela 97 - Matriz de Confusão para Redes Neurais (2 camadas com 20 neurônios cada) em BD3 com 4 classes

	<i>Excelente (predita)</i>	<i>Boa (predita)</i>	<i>Média (predita)</i>	<i>Ruim (predita)</i>
<i>Excelente (real)</i>	106	23	6	2
<i>Boa (real)</i>	34	13	3	4
<i>Média (real)</i>	14	3	2	2
<i>Ruim (real)</i>	7	7	2	12