

Classificação de falhas de equipamentos de unidade de intervenção em construção de poços marítimos por meio de mineração textual

Marcus V. G. Cesar¹, Marley M. B. R. Vellasco², Karla Figueiredo³

¹Petrobras – Petróleo Brasileiro S.A.

Caixa Postal 11010-110 – Santos – SP - Brazil

²Departamento de Engenharia Elétrica - PUC-RIO – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) – Rio de Janeiro – RJ –Brazil

³Departamento de Ciências da Computação - UERJ – Universidade do Estado do Rio de Janeiro (UERJ) – Rio de Janeiro– RJ –Brazil

marcusgiollo@petrobras.com.br, marley@ele.puc-rio.com.br,
karlafigueiredo@ime.uerj.br

Abstract. *This paper proposes the development of a tool to optimize classification of textual records related to failures in offshore drilling rig equipment. Given the high cost of offshore drilling rigs and the high amount of time spent reading and interpreting textual data, this tool translates into a practical and efficient way to tackle this type of data. After some adjustments on the database, the tool presented better results than those obtained on the evaluation phase.*

Resumo. *Este trabalho propõe o desenvolvimento de uma ferramenta para otimizar a classificação de registros textuais de falhas ocorridas em equipamentos de plataformas de perfuração de poços marítimos. Dado o alto custo de utilização destas plataformas e o elevado tempo dispendido para leitura e interpretação de dados textuais, tal ferramenta se traduz em uma forma prática e eficiente de tratamento deste tipo de dado. Após correções na base de dados, a ferramenta apresentou resultados acima dos obtidos em fase de avaliação.*

1. Introdução

A construção de poços marítimos tem se mostrado uma atividade complexa e de alto risco [Kaiser, 2009]. Para efetuar esta atividade as empresas operadoras se valem principalmente das unidades de intervenção de poços, também conhecidas como sondas marítimas. Elas representam alto custo diário na construção dos poços [IHS, 2016] e por este motivo recebem o maior foco de atenção no controle de suas atividades. As falhas nos seus equipamentos se traduzem em perdas significativas tanto de eficiência operacional bem como monetárias. Em junho de 2011, foi implantado no banco de dados da Petrobras (baseado no software *Open Wells*) [Landmark, 2017] um maior detalhamento na classificação das falhas de equipamentos de sonda. Com isso gerou-se uma descontinuidade nos registros da empresa em relação aos dados anteriores a 2011, menos detalhados. Assim, surgiu a demanda para adequar estes casos menos detalhados à classificação atual, mais completa.

Os registros são compostos basicamente de informação textual. Para um passivo de 3384 registros de intervenções de perfuração de poços, demandar-se-ia muito tempo para um operador realizar a tarefa de ler e classificar os registros um a um. Com isso vislumbrou-se uma ferramenta, denominada TEXTCLASS-BR, que pudesse classificar de forma mais automatizada possível este passivo, utilizando os registros feitos após junho de 2011 (já classificados) como base de treinamento do algoritmo.

Assim, os principais objetivos deste trabalho são: sanar esta descontinuidade nos registros de falha de equipamentos de sonda; sugerir melhorias nos registros das falhas de equipamentos de sonda, caso necessário; e fomentar o uso de Inteligência Artificial aplicada à classificação de documentos textuais no cenário de engenharia de poços.

Este trabalho está organizado em mais 4 seções. A Seção 2 aborda a contextualização do problema no cenário de construção de poços marítimos da Petrobras. A Seção 3 detalha como se desenvolveu o modelo de categorização textual. A Seção 4 aborda a aplicação do modelo, através de estudo de casos. Nesta seção também são destacadas as correções efetuadas na base de dados. A seção 5 apresenta as conclusões finais.

2. Falhas de equipamentos de unidades de intervenção de poços marítimos

Este trabalho se concentrou nas unidades de intervenção de poços marítimos, também conhecidas como sondas flutuantes. Estas unidades possuem vários sistemas complexos e integrados. No software de relatórios de intervenções em poços *Open Wells* (atualmente empregado na área de engenharia de poços da Petrobras para registro das intervenções), estes sistemas são classificados conforme indicado na Figura 1.

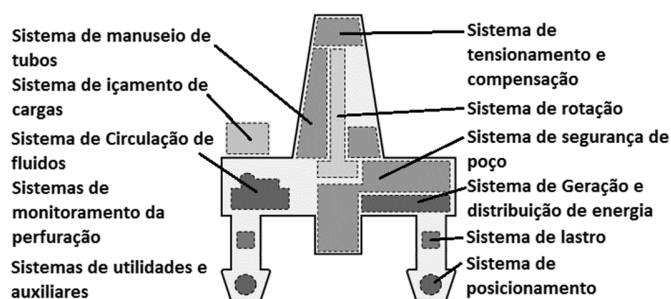


Figura 1: Sistemas existentes em sondas marítimas.

O novo detalhamento das anormalidades, instituído em 2011, divide as sondas marítimas em 11 sistemas, ou classes. Este detalhamento, manifestado nessas novas classes, foi instituído somente em 2011. Antes disso, todos os sistemas pertenciam a uma mesma classe denominada “equipamentos de sonda”. O advento das novas classes criou uma descontinuidade na base de dados. Para os registros das intervenções de perfuração (foco deste trabalho), de 2008 até 2011 gerou-se um passivo de 3384 registros que deveria ser adequado às novas classes criadas.

Para o desenvolvimento do modelo, foram coletados 9202 registros (já rotulados com as novas classes), contemplando o período de 2011 até 2013. A distribuição das classes desses novos registros é pouco balanceada, com as quatro maiores classes representando mais

de 80% da base de dados, enquanto outras correspondendo a menos de 1% dos registros, conforme apresentado na tabela 1.

Tabela 1: Distribuição das anormalidades de falha de equipamentos de sonda, a serem utilizadas para modelagem do classificador.

Classe	Frequência	Fração da base
1	3077	33,44%
2	1607	17,46%
3	1576	17,13%
4	1522	16,54%
5	432	4,69%
6	266	2,89%
7	209	2,27%
8	187	2,03%
9	172	1,87%
10	142	1,54%
11	12	0,13%

Os dados utilizados neste trabalho consistem basicamente em informações inseridas pelos operadores relacionadas às falhas de equipamento de sonda, dentre elas: título resumido da falha ocorrida; campo de descrição detalhada do evento; datas de início e fim da interrupção das atividades da unidade de intervenção. Deste modo, há 2 campos de informação puramente textual (título e descrição) e um campo numérico, resultante da diferença entre o momento de início e fim da ocorrência da falha. A Figura 2 ilustra a disposição destes campos no software de registro das falhas de equipamento após a implementação do detalhamento ocorrido em 2011.

Título	Falha no sistema tensionador de riser
Descrição	Efetuada troca dos discos de ruptura dos tensionadores 5, 6, 7, 10, 11 e 13. Calibrado nível de fluido em todos os tensionadores. Calibrado sistema tensionador de riser.
Data/hora de Início	14/12/2013 01:00
Data/hora do Fim	14/12/2013 08:00

Figura 2: A esquerda como os dados estão dispostos no software Open Wells. À direita, um resumo das informações coletadas.

Nos registros efetuados antes de 2011 não havia detalhamento nível 2, de forma que todas as falhas pertenciam à mesma classe. O objetivo do modelo é classificar o passivo de registros dentro das 11 classes implementadas.

3. Sistema TEXTCLASS-BR

O sistema TEXTCLASS-BR segue o processo de mineração textual composto de: Coleta de dados, Pré-processamento, Indexação, Balanceamento e Classificação [Aranha, 2009]. Como as classes eram altamente desbalanceadas, foram testados algoritmos de balanceamento com

o intuito de aumentar o desempenho do sistema. Foram analisadas várias configurações do sistema TEXTCLASS-BR, até obtenção da configuração campeã.

Foi aplicado o modelo de representação de espaço vetorial [Silva, 2007] por meio da MTD (matriz de termos por documentos), que é um produto da terceira etapa (indexação). Ela visa consolidar todos os documentos textuais em uma matriz de dados numéricos. As seções a seguir descrevem em detalhes o sistema TEXTCLASS-BR.

3.1 Coleta

Inicialmente houve a coleta de dados textuais e numéricos na base de dados de construção de poços marítimos da Petrobras por meio de *queries*. Foram selecionados todos os eventos de falha de equipamentos de sonda que ocorreram nas intervenções de perfuração, de 2008 a 2013. Conforme apontado na seção 2, utilizaram-se os dados textuais das anormalidades, compostos de 2 campos: título e descrição detalhada da falha ocorrida. Já para os dados numéricos, coletaram-se os momentos de início e fim de ocorrência da falha.

3.2 Pré-processamento

O pré-processamento foi efetuado para os dados textuais e numéricos. Para o tratamento dos dados textuais, o objetivo principal é aumentar o desempenho do algoritmo por meio da seleção e consolidação de termos textuais mais relevantes e exclusão de termos textuais menos importantes.

Com este intuito, foram desempenhadas nesta etapa as seguintes atividades: criação de uma lista de termos múltiplos, que consiste em concatenar palavras compostas (por exemplo “ferro de passar”); remoção de termos de pouca relevância, como termos que possuem baixíssima frequência, além de termos numéricos; descaracterização de nomes próprios (case folding) conforme [Lopes, 2004]; remoção de acentuação [Dias & Malheiros, 2004]. Finalmente, foi aplicado do algoritmo de Porter, conforme [Paice, 1983], para efetuar a normalização morfológica dos termos (em inglês, stemming).

Para os dados numéricos relativos à duração das anormalidades, efetuou-se o cálculo da duração da falha, por meio da diferença entre sua data/hora de início e fim.

3.3 Indexação

Na etapa de indexação dos dados textuais, após a redução dos termos na etapa de pré-processamento, cria-se um dicionário de sinônimos. Ele consolida termos de mesmo significado em um só termo principal.

Conforme apontado no início da seção 3, após todo o tratamento dos termos da base de dados, a MTD pode ser criada. A métrica mais intuitiva utilizada para dimensionar cada termo é a frequência simples dos termos em cada respectivo documento. Este seria um tipo de métrica definidora da importância de um termo [Gomes, 2013]. Para este trabalho foram avaliadas as seguintes métricas (também referenciadas como “pesos”): Term Frequency (TF): frequência dos termos no documento; Booleano (B): presença (1) ou não (0) do termo no documento; e Term frequency – inverse document frequency (TF-IDF): peso do termo considerando sua frequência local e global, conforme [Salton, McGill, 1986].

Visando reduzir a quantidade de termos da MTD foi avaliada uma ferramenta de remoção de termos mais esparsos na MTD, conforme especificado em [Feinerer, 2013]. Pode-se definir numericamente a esparsidade de um termo como o percentual de documentos onde não há presença do termo. Por exemplo, um termo com esparsidade 90% aponta que este termo está presente em 10% dos documentos de toda coleção. Com isso, aplicando-se um limite de corte, pode-se retirar da MTD termos demasiadamente raros. Foram avaliados três níveis de redução de esparsidade, conforme [Feinerer, 2015]: sem redução; redução moderada; e redução agressiva.

Para finalizar a etapa de indexação, os dados devem ser normalizados. De acordo com [Hsu, Chang, Lin, 2010], este artifício aumenta substancialmente o desempenho em algoritmos de aprendizado de máquina (particularmente em Support Vector Machines - SVMs). O intervalo para normalização dos dados da MTD utilizado foi de [0,1], conforme sugerido na mesma referência.

Já a variável relacionada à duração das falhas não foi submetida ao método de normalização simples por extremos, devido à grande amplitude de valores para este tipo de dado, quando comparado com os valores gerados nas MTDs. A duração da maioria das anormalidades (em torno de 80%) está contida em um intervalo de duração de até 12 horas, enquanto anormalidades mais demoradas podem chegar a valores acima de 600 horas. Com isso optou-se por normalizar os valores desta variável em 2 etapas: de zero até 0,8 as anormalidades com duração até 12 horas, e as demais seriam normalizadas de 0,8 até 1.

3.4 Balanceamento

Conforme especificado na seção 2, devido à distribuição de frequência não uniforme entre as classes analisadas, valeu-se de ferramentas de balanceamento de dados conforme apontado em [He, Garcia, 2009]. Foram utilizados métodos de aumento de classes minoritárias e de redução de classes majoritárias indicados nesta referência. Dentre os métodos de redução de dados, optou-se pela redução de dados de forma aleatória, Tomek Links, ENN e NCL [He, Garcia, 2009]. Já para o caso de balanceamento para aumento de classes minoritárias, utilizou-se o método de geração aleatória de dados e o SMOTE [He, Garcia, 2009]. Todas as etapas da modelagem, até o balanceamento da base de dados, são resumidas no diagrama da Figura 3.

3.5 Classificação

O Módulo de Classificação foi implementado, neste caso específico, por meio de *Support Vector Machines* – SVM para classificação, denominada *Support Vector Classification* – SVC. O diagrama da Figura 4 ilustra de forma resumida esta etapa do processo:

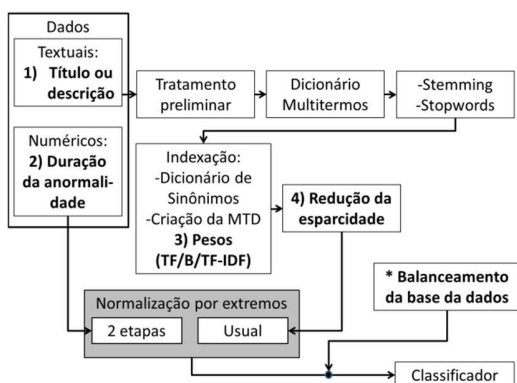


Figura 3: Etapa de pré-processamento.

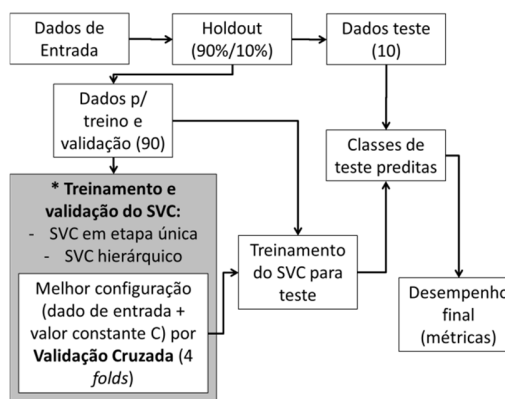


Figura 4: Etapa de classificação dos documentos.

Os dados são primeiramente divididos em 2 frações iniciais: 90% para dados de treinamento e validação e 10% são separados para o teste final, técnica denominada holdout [Soares, 2013]. O treinamento e validação da SVC é feito por meio da técnica de validação cruzada [Soares, 2013], sendo utilizados, neste trabalho, 4 conjuntos (folds). Após treinamento e validação da SVC, a mesma é submetida ao teste final, que é feito sobre os 10% separados inicialmente. Em seguida são obtidas as configurações com melhor desempenho para os vários estudos de caso.

O treinamento da SVC foi realizado de duas formas: em etapa única (abordagem mais usual); e de forma hierárquica, dividindo o problema em duas etapas, conforme exemplificado na Figura 5.

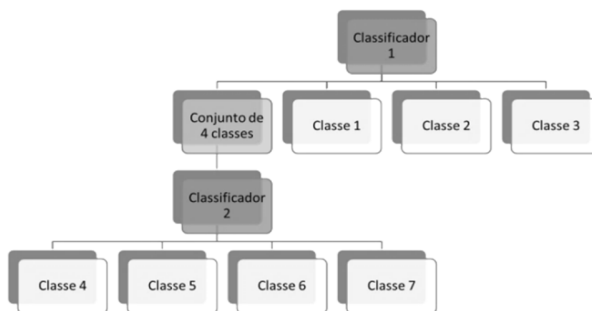


Figura 5: Método de classificação hierárquico [Cesar, 2017].

No caso do método de classificação hierárquico o resultado de aplicação de um primeiro conjunto de SVCs é enviado e submetido a um segundo conjunto de SVCs, treinado de forma independente do primeiro. Conforme apontado na Figura 5, a ideia é especializar o classificador 1, reduzindo o número de classes as quais ele é submetido. A mesma ideia se aplicaria ao classificador 2. Desta forma, há 2 configurações possíveis no que tange à quantidade de etapas para classificação: etapa única e hierárquico (2 etapas).

Com relação aos SVCs aplicados em classificação, as seguintes abordagens foram utilizadas neste trabalho: uso de variáveis soltas (slack) [Gunn, 1998], por meio de constantes de regularização (parâmetro C), utilizando-se da heurística proposta em [Hsu, Chang, Lin, 2010] para obtenção deste parâmetro. Baseado nesta mesma referência, foi utilizada a função

Kernel linear que é efetiva em aplicações de classificação textual e para tratar o problema de várias classes com uso de SVCs, valeu-se da abordagem “um contra um”, combinando os resultados por meio da estratégia de votação [Hsu, Lin, 2012].

Todas as configurações dos modelos nos vários estudos de caso foram, inicialmente, avaliadas através das métricas para várias classes, dispostas em [Sokolova, Lapalme, 2009]: Macro F-score e Micro F-score. Entretanto, a Macro F-score não apresentou resultados satisfatórios por limitações do número de registros em classes menores [Cesar, 2017]. Assim, utilizou-se como métrica principal o Micro F-score, o qual privilegia classes de maior número. Como alternativa ao Macro F-score, utilizou-se a métrica de média aritmética simples do percentual de acerto das classes como uma referência auxiliar.

Esta etapa pode ser resumida no seguinte diagrama:

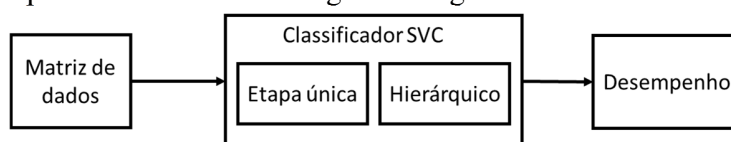


Figura 6: Diagrama do modelo de classificação utilizado.

Ao final, houve um total de 36 configurações básicas que foram analisadas em todos os estudos de caso, geradas pela combinação das possíveis variações: com relação ao dado textual utilizado (título ou descrição das falhas); peso utilizado na geração da MTD; redução da esparsidade; utilizar ou não o dado relativo a duração das falhas.

4. Estudos de caso

De forma a se obter uma referência inicial para efeitos de comparação (*baseline*), especialistas avaliaram 461 registros de falha gerados pelos operadores. O intuito desta avaliação é aferir a precisão de preenchimento dos registros pelos operadores, de modo a comparar com o desempenho obtido pelos modelos em análise nos estudos de caso. Os resultados foram os seguintes: Micro F-score 35,63% e Média das classes de 67,59%. Foram observados erros de preenchimento por parte dos operadores, mostrando que a base de dados poderia conter uma parcela significativa de registros enviesados.

Assim, foram realizados 9 estudos de casos, divididos em três etapas: sem correção da base de dados; após primeira correção; após segunda correção. Em cada estudo de caso foram avaliadas as 36 configurações básicas do modelo, além de alterações em sua estrutura, de forma a adequá-lo progressivamente a partir dos resultados obtidos em cada estudo de caso. Esta evolução pode ser observada na Tabela 2.

O estudo de caso inicial avaliou todas as configurações básicas para o SVC de etapa única, sem balanceamento. Decidiu-se por testar a classificação hierárquica separando na primeira etapa as classes 1 a 4 (mais numerosas), e na segunda etapa avaliando as demais classes. Foram aplicados os métodos de balanceamento descritos na seção 3.4 da seguinte forma: aplicou-se a redução de dados na classe 1, pois ela era a classe mais numerosa (33,44% da base de dados) e efetuou-se o aumento das classes de 6 a 11, pois estas possuíam tamanho reduzido. Para a redução da classe o método NCL se sobressaiu aos demais e para o aumento de classes o método SMOTE obteve melhor desempenho. Após os primeiros sete estudos de

caso a Tabela 3 foi elaborada, trazendo um resumo do desempenho para as métricas Micro F-score e Média de acerto das classes. A nomenclatura “11C” se refere ao SVC de etapa única e a nomenclatura “4+7C” se refere ao SVC de 2 etapas (classificação hierárquica).

Tabela 2: Resumo dos estudos de caso abordados.

Estudo de caso	Correção na base de dados	Descrição
1	Sem correção	Inicial
2		Classificação hierárquica (4+7)
3		Redução da classe 1
4		Estudos 2 e 3
5		Aumento das classes 6 a 11
6		Estudos 2 e 5
7		Estudos 1, 2, 3 e 5
8	Após 1ª correção	Classificação hierárquica (5+6) e aumento das classes 10 e 11
9	Após 2ª correção	Revisão dos estudos de 1 a 8

Os desempenhos dos três melhores modelos foram destacados em negrito. Na métrica Micro F-Score o SVC de 11C sem balanceamento obteve o melhor resultado (43,18%). O segundo melhor modelo nesta mesma métrica, que seria o SVC de 11C com aumento das classes menores por SMOTE, teve desempenho ligeiramente inferior ao melhor modelo (43,01%), porém este se sobressaiu na métrica Média das classes, com resultado de 69,14% (o modelo de SVC de 11C sem balanceamento obteve 68,00% de desempenho nesta métrica). Como a métrica principal atribuída foi Micro F-score, o SVC de 11C sem balanceamento foi considerado o melhor modelo obtido até então.

Tabela 3: melhores resultados obtidos sem correção da base de dados.

		Balanceamento			Sem balanceamento
		Redução classe 1	Aumento das classes menores	Ambos balanc.	
			NCL	SMOTE	
Micro F-score	SVC 11C	39,47%	43,01%	40,56%	43,18%
	SVC 4+7C	32,11%	38,14%	28,21%	40,56%
Média das classes	SVC 11C	60,35%	69,14%	63,62%	68,00%
	SVC 4+7C	61,32%	60,51%	57,12%	63,47%

Para as configurações básicas descritas ao final da seção 3.5, vale destacar que os classificadores que utilizaram os dados do campo do título das falhas, bem como o campo de duração das falhas se destacaram nos estudos de caso.

Conforme descrito no início desta seção, a avaliação do desempenho dos operadores pelos especialistas mostrou que a base de dados possuía uma fração não desprezível de apontamentos equivocados. Dada essa suspeita, os erros de apontamento do melhor modelo na classificação da base de teste foram avaliados, com intuito de verificar se os erros eram do modelo ou se havia erros de classificação por parte dos operadores. Após confirmar erros sistemáticos por parte dos operadores na base de testes, foram avaliados pontualmente erros

semelhantes na base de dados de treinamento (mais extensa). Na base de dados de treino foram corrigidos 321 documentos (3,8% de 8290 documentos), já na base de teste foram corrigidos 106 documentos (11,5% de 922 documentos).

Após esta primeira correção, foi iniciado o estudo de caso 8, que basicamente repetiu o estudo de caso 7, com algumas adequações. Inicialmente o melhor modelo da etapa anterior foi avaliado, para verificar o impacto da correção. O desempenho do classificador nesta avaliação apresentou sensível melhora: Micro F-score 61,84% e Média das classes de 80,34, com uma melhora no desempenho para as classes 5 e 6. Este fato motivou uma alteração na abordagem classificador hierárquico: antes o classificador hierárquico “4+7” abarcava na primeira etapa as classes de 1 a 4 (mais numerosas), e na segunda etapa as 7 classes remanescentes, de 5 a 11. Como esta abordagem não trouxe bons resultados, nesta nova proposta, a primeira etapa incluirá também as classes 5 e 6, de forma que as seis classes (de 1 a 6), que tiveram alta taxa de acerto pelo modelo, serão classificadas na primeira etapa e as cinco demais classes (7 a 11) serão classificadas na segunda etapa. Este modelo foi denominado hierárquico “6+5”.

Além disso, após se avaliar a eficácia no balanceamento da base de dados até então, se propôs o aumento somente para as classes 10 e 11, que tiveram baixo desempenho do classificador bem como possuem poucas observações na fase de treinamento e validação (1,54% da base para a classe 10 e 0,13% dos elementos da base para a classe 11). A redução da classe 1 foi suprimida nesta nova etapa, pois prejudicou o desempenho em todos os estudos de casos. O aumento das classes 6, 7, 8 e 9 também foi suprimido, pois o aumento dessas classes não trouxe ganho relevante. A Tabela 4 apresenta um resumo dos resultados consolidados do estudo de caso 8.

O melhor modelo na métrica micro F-score foi o modelo SVC em 2 etapas (6-5) sem balanceamento. Primeiramente observa-se que as novas propostas sugeridas para este estudo de caso, sejam elas: a classificação em duas etapas e o balanceamento somente das classes 10 e 11, melhoraram o desempenho do modelo, quando se compara o melhor modelo deste estudo de caso com o melhor modelo obtido anteriormente (estudo de caso 1). No estudo de caso 8 os modelos que utilizaram os dados do campo do título das falhas, bem como o campo de duração das falhas tiveram desempenho superior aos demais.

Devido à acentuada melhoria no desempenho do modelo, decidiu-se efetuar uma última correção, mais abrangente que a primeira. Para esta correção foi utilizado o melhor modelo até então encontrado pela métrica Micro F-score: classificação hierárquica (2 etapas – “6+5”), sem balanceamento da base de dados.

Todos os erros do classificador nesta configuração foram analisados, totalizando 2047 registros. Observou-se que os operadores confundiram bastante as classes 1, 2, 4 e 7, especialmente a classe 1. Após esta segunda correção, que ocorreu nas bases de treinamento e teste, efetuou-se a última avaliação de configurações do modelo, no estudo de caso 9. Seguiu-se o mesmo formato do estudo de caso 8, em termos de configurações, cujos resultados constam na tabela 5.

Um ponto interessante observado foi que o desempenho do algoritmo caiu da primeira para a segunda correção (apesar de se esperar um aumento do mesmo). Este fato pode ser

explicado pelo método utilizado para efetuar as correções. Para uma correção mais eficiente, foi utilizada a heurística de avaliar as classes atribuídas de forma errada pelo classificador. Porém em momento algum foram conferidas as observações onde o classificador acertou. Dada a taxa de erro do operador, o que garante que ele pode ter atribuído classes erradas e o classificador ter aprendido esta atribuição também de forma errada? Além disso, quando se efetuou a primeira correção, onde se atuou somente na base de teste, a avaliação dos modelos teve confiabilidade aumentada, mas a base de treinamento ainda carecia de correções. Após a segunda correção, onde o foco maior foi na base de treinamento, apesar da queda no desempenho, pode-se inferir que a confiabilidade dos resultados aumentou.

Tabela 4: Desempenho das melhores configurações para o estudo de caso 8.

Melhores resultados (Micro F-score)	Aumento 10 e 11	Sem balanceamento
Modelo SVC		
Etapa única	64,16%	61,84%
2 etapas (6-5)	53,00%	65,07%
Melhores resultados (Média das classes)	Aumento 10 e 11	Sem balanceamento
Modelo SVC		
Etapa única	80,26%	80,34%
2 etapas (6-5)	75,25%	79,27%

Tabela 5: Desempenho das melhores configurações para o estudo de caso 9.

Melhores resultados (Micro F-score)	Aumento 10 e 11	Sem balanceamento
Modelo SVC		
Etapa única	62,13%	61,84%
2 etapas (6-5)	57,05%	62,99%
Melhores resultados (Média das classes)	Aumento 10 e 11	Sem balanceamento
Modelo SVC		
Etapa única	79,89%	80,66%
2 etapas (6-5)	77,14%	79,89%

Após a obtenção do melhor modelo a ser utilizado no sistema TEXTCLASS-BR, efetuou-se o processo descrito na Figura 10 para classificar o passivo de anormalidades. Conforme apontado nesta figura, consolidou-se a base de treinamento, validação e teste (ou seja, conjunção de todos os dados) para treinar o sistema. Após esta etapa, classificou-se o passivo de anormalidades.

Com o resultado da classificação gerada pelo algoritmo de melhor configuração, foram selecionadas aleatoriamente e classificadas 208 falhas (do passivo de 3384 falhas) por um especialista. O objetivo desta etapa foi de avaliar a eficácia de classificação do melhor modelo. Os resultados foram: métrica Micro F-score de 70,05% e média de acerto das classes em 94,27%. O sistema classificou esta amostra de 208 falhas do passivo com alta taxa de acerto, inclusive com desempenho superior à fase de obtenção do melhor modelo. Comparando este resultado com o melhor resultado na fase de teste final e escolha do melhor modelo (Tabela 5), observa-se que o valor de Micro F-score é 7,06 pontos percentuais maior,

e a média de acerto entre classes é 14,38 pontos percentuais maior, mostrando alto desempenho do sistema TEXTCLASS-BR.

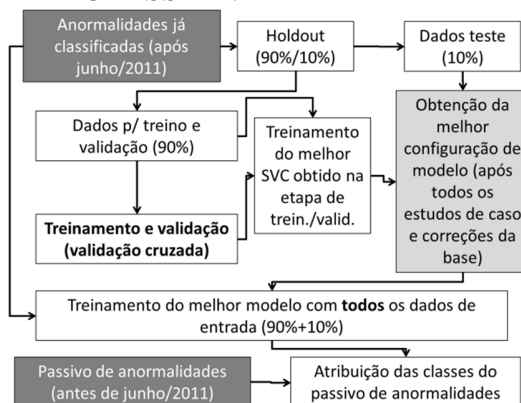


Figura 7: Diagrama para correção do passivo dos registros.

5. Discussão

Este trabalho teve como objetivo principal desenvolver um classificador de falhas de equipamentos de sonda para a Petrobras. O TEXTCLASS-BR atualizou o passivo de registros relacionados às falhas de equipamentos de unidades de intervenção ocorridas em intervenções de perfuração em sondas marítimas preenchidas do início de 2008 até junho de 2011, com base em uma nova classificação que foi aplicada a partir deste ano.

Para alcançar este objetivo, foram necessárias duas correções na base de dados. Ao efetuar a avaliação final do desempenho do modelo ao classificar o passivo dos registros de falha de equipamentos, observou-se que o melhor modelo superou as expectativas, com alto índice de acerto tanto para a métrica micro F-score (70,05%) quanto para a taxa média de acerto entre classes (94,27%).

A correção na base de dados colaborou de forma a orientar os operadores quanto ao preenchimento classificação da classe 4, muito confundida com a classe 1. Além disso, trouxe luz à necessidade de suprimir algumas classes de tamanhos muito pequenos (como as classes 10 e 11), de modo a inseri-las em outras classes. O mesmo se aplica à numerosa classe 1, que poderia ser dividida em classes menores. Isto traria mais uniformidade, não só em análises de mineração de dados, mas na análise das falhas de equipamentos pelo corpo gerencial da empresa. Para um maior detalhamento deste trabalho, consultar [Cesar, 2017].

Referências

- Aranha, C. N. (2007) Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional, Tese de Doutorado, PUC-Rio, Rio de Janeiro, Brasil.
- Cesar, M. V. G. (2017) Classificação de falhas de equipamentos de unidades de intervenção em construção de poços marítimos por meio de mineração textual. Dissertação de mestrado, PUC-Rio, Rio de Janeiro, Brasil.

- Cortes C., Vapnik V. (1995) Support Vector Networks. *Machine Learning* 20:273-297.
- Dias, M. A. L., Malheiros M. G. (2004) Estudo de Técnicas de Radicalização para a Língua Portuguesa. Centro Universitário UNIVATES, Lajeado, Rio Grande do Sul, Brasil.
- Feinerer I., Hornik K. (2015) Tm: Text Mining Package. R package version 0.6-2.
- Gunn, S. (1998) Support Vector Machines for Classification and Regression. ISIS Technical Report.
- Gomes, N. O. (2013) Categorização de textos - estudo de caso: documentos de pedidos de patente no idioma português. Tese de Doutorado, PUC-Rio, Rio de Janeiro, Brasil.
- He H., Garcia E. A. (2009) Learning from Imbalanced Data. *IEEE Trans. Knowledge and Data Engineering*, vol. 21, issue 9, pp. 1263-1284.
- Hsu C., Chang C., Lin C. A (2010) Practical Guide to Support Vector Classification. Department of Computer Science, National Taiwan University, Taipei, Taiwan.
- Hsu, C., Lin, C. (2012) A Comparison on Methods for Multi-class Support Vector Machines. In *IEEE Transactions on Neural Networks*, vol. 13 (2), pp. 415-425.
- IHS, (Information Handling Services Petrodata) (2016) Offshore Rig Day Rate Trends. Disponível em: <https://www.ihs.com/products/oil-gas-drilling-rigs-offshore-day-rates.html>. Acessado em: 01/02/2017.
- Kaiser, M. J. (2009) Modeling the time and cost to drill an offshore well. *Elsevier Energy Journal*, vol. 34, Issue 9, pp. 1043-1422.
- Landmark (2017) Open Wells® Operations Reporting – Software para registro de operações em construções de poços. Disponível em: < <https://www.landmark.solutions/OpenWells>>. Acesso em: 12 de jan. 2017.
- Lopes, M. C. S. (2004) Mineração de dados textuais utilizando técnicas de clustering, para o idioma português. Dissertação de mestrado, UFRJ, Rio de Janeiro, Brasil.
- Paice, C. D. (1983) Another Stemmer. Department of Computing. Lancaster University, Reino Unido.
- Salton G., Mcgill M. J. (1986) Introduction to modern information retrieval. McGraw-Hill.
- Silva, A. A. (2007). Aûuri: Um portal para Mineração de Textos integrado a Grids, Dissertação de Mestrado, UFRJ, Rio de Janeiro, Brasil.
- Soares, F. A. (2013) Categorização Automática de Textos Baseada em Mineração de Textos, Tese de Doutorado, PUC-Rio, Rio de Janeiro, Brasil.
- Sokolova M., Lapalme, G. (2009) A systematic analysis of performance measures for classification tasks. *Information Processing and Management* 45, pp. 427-437. Elsevier.