

# Análise automática de relatórios de acreditação usando aprendizagem de máquina

Oduvaldo Vick Neto<sup>1</sup>, Luiz F. R. C. Carmo<sup>1</sup>, Rosembergue P. Souza<sup>1</sup>

<sup>1</sup>Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro)  
Av N. S. das Graças, 50, 25.250-020, Xerém – Duque de Caxias – Rio de Janeiro

{oneto, rpereira, lfrust}@inmetro.gov.br

**Abstract.** *During an accreditation process of inspection bodies, reports need to be reviewed by a technical team from Inmetro. Using the answers given in these reports, this study proposes a procedure for the automatic analysis of Inmetro accreditation reports. The objective is to categorize the report components into adequate or inappropriate, avoiding the need for manual review. Furthermore, it is intended to increase the efficiency of the models using customized features, which can be defined as important elements for an answer to be considered adequate. In the experiments, the SVM was the algorithm with the best result for the problem and the use of customized features improved the final performance depending on the question and the algorithm used.*

**Resumo.** *Durante um processo de acreditação de organismos de inspeção, relatórios precisam ser revisados por uma equipe técnica do Inmetro. Utilizando as respostas dadas nesses relatórios, este estudo propõe um procedimento para análise automática dos relatórios de acreditação do Inmetro. O objetivo é categorizar os componentes dos relatórios em adequados ou inadequados, evitando a necessidade de revisão manual. Além disso, pretende-se aumentar a eficiência dos modelos utilizando features customizadas, que seriam elementos identificados como importantes para que uma resposta seja considerada adequada. Nos experimentos, o SVM foi o algoritmo com melhor resultado para o problema e a utilização de features customizadas melhorou o desempenho final a depender da pergunta e do algoritmo utilizado.*

## 1. Introdução

Acreditação é o reconhecimento formal da competência dos Organismos de Avaliação da Conformidade (OAC) para atenderem requisitos previamente definidos e realizar suas atividades com confiança [INMETRO 2021]. Ela é uma ferramenta utilizada internacionalmente com o objetivo de gerar confiança nas organizações que atuam em diferentes tipos de esquemas de acreditação. Dentre os diversos esquemas realizados pelo Inmetro, existe a acreditação de organismos de inspeção. O esquema de inspeção no Inmetro engloba as áreas de transporte de produtos perigosos, segurança veicular, ensaios não destrutivos, eficiência energética de edificações e instalações prediais de gás combustível. O principal foco desses esquemas de inspeção é a segurança e desempenho, buscando reduzir os riscos de compradores, proprietários e usuários nas áreas citadas.

No processo de acreditação dos organismos de inspeção, uma das etapas consiste em formar uma equipe avaliadora que irá analisar a documentação do organismo

solicitante e que deve responder uma série de perguntas de um relatório a respeito do organismo. O relatório possui 419 perguntas e posteriormente as respostas precisam ser revisadas por uma equipe técnica do Inmetro.

Este trabalho pretende trazer duas contribuições, sendo a primeira a proposta de um procedimento para análise automática de relatórios de acreditação utilizados pelo Inmetro. Por sua vez, a segunda contribuição consiste na aplicação de *features* customizadas para melhoria do desempenho dos algoritmos na obtenção de modelos capazes de analisar os relatórios de acreditação.

## 2. Trabalhos Relacionados

Existem diversos trabalhos na literatura sobre classificação de texto e sua aplicação. Serão apresentados alguns que possuem semelhanças com o trabalho proposto em relação aos algoritmos e técnicas utilizadas.

Em [Bahgat et al. 2018] é proposta uma nova forma de classificação binária de e-mails para identificar quais são e quais não são *spam*. No pré-processamento são eliminados números e símbolos, é realizada a tokenização em TF-IDF e uma redução semântica de *features* (o diferencial do trabalho), com posterior seleção de *features* através de PCA e de CFS (seleção de *features* correlacionadas). Os algoritmos de classificação utilizados na comparação são NB, SVM, LR, J48, RF e a rede neural de função de base radial. O objetivo do trabalho era indicar que a redução semântica de *features* e a utilização de CFS melhoram ou mantém o desempenho dos algoritmos na classificação de e-mail. Os melhores resultados foram obtidos pelos algoritmos SVM, NB e LR, com *F1-Score* acima de 90%.

Em [Aletras et al. 2016] a classificação de texto é utilizada para prever decisões judiciais na corte europeia de direitos humanos, onde através dos textos dos julgamentos o classificador pretende realizar a classificação binária entre “houve uma violação de um artigo da convenção de direitos humanos” e “não houve uma violação de um artigo da convenção de direitos humanos”. No estudo foram selecionados 3 artigos da convenção de direitos humanos para que fosse verificada sua violação, o algoritmo utilizado foi o SVM e a tokenização através de *Bag-of-Words* (BOW). A solução foi testada em cada tópico individual do julgamento e também no texto completo, obtendo uma acurácia de mais de 70%.

Em [Caccamisi et al. 2020] é feita uma classificação dos pacientes através de relatórios médicos eletrônicos pré-existentes em 3 categorias: fumantes, não fumantes e ex-fumantes. Foram utilizados os algoritmos SVM, NB, KNN e J48 com um pré-processamento que consistiu na aplicação de *lowercase* e tokenização em BOW com apenas unigrams e também de forma híbrida com 1-gram e 2-gram. O resultado com melhor performance foi o SVM com tokenização híbrida, chegando a um *F1-Score* de 98,1%.

Em [Lei et al. 2017] é feita uma classificação de julgamentos em 13 categorias relacionadas à qualidade de produtos, com base no padrão legal da divisão da indústria chinesa. Na etapa de pré-processamento foram removidas as *stopwords* não apenas da língua chinesa, mas também foram identificadas palavras do meio jurídico que estão sempre presentes e tem pouco significado, além disso foi realizada a tokenização em TF-IDF e utilizadas 3 técnicas de redução de *features*: mínima frequência de documentos,

PCA e SVD. Os algoritmos utilizados foram o NB, a árvore de decisão, a RF e o SVM. O resultado com melhor resultado foi o SVM com um *F1-score* de 87%.

O trabalho proposto também fará uma classificação binária assim como [Bahgat et al. 2018] e [Aletras et al. 2016], e utilizará a tokenização em TF-IDF como [Bahgat et al. 2018] e [Lei et al. 2017], visto que o TF-IDF também considera a frequência de documentos em que o token aparece ao invés de apenas contabilizar a quantidade de ocorrências total do mesmo como ocorre no BOW. Será utilizado o modelo híbrido de tokenização em 1-gram e 2-gram assim como no modelo de melhor resultado obtido por [Caccamisi et al. 2020]. Não será aplicado nenhum método de redução ou seleção de *features* como ocorreu em [Bahgat et al. 2018] e [Lei et al. 2017], visto que os algoritmos de classificação escolhidos (NB, SVM, LR e RF) possuem um bom desempenho para altas dimensões [Caruana et al. 2008] [Palatucci and Mitchell 2007] [Ukey and Alvi 2012]. A remoção de *stopwords* corriqueiras do meio jurídico feita em [Lei et al. 2017] pode ser considerada o inverso do que está sendo proposto com as *features* customizadas, pois ao invés de detectar as palavras menos importantes para serem retiradas são detectadas as mais importantes para elevar seu peso no modelo, algo que não é feito por nenhum dos estudos apresentados e que é um diferencial deste trabalho.

### 3. Proposta

A Figura 1 mostra o fluxograma utilizado para criar o mecanismo de análise automática de relatórios de acreditação do Inmetro. As etapas do fluxograma serão detalhadas nas subseções seguintes.

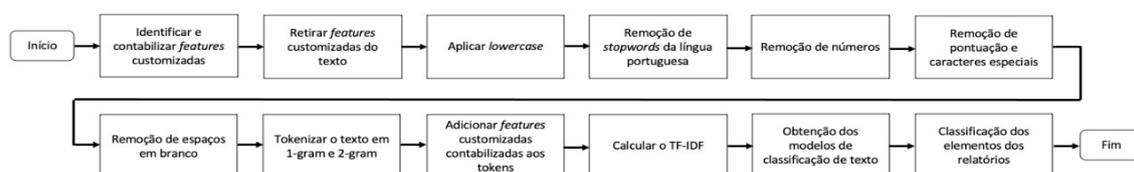


Figura 1. Fluxograma de criação do mecanismo de análise automática de relatórios de acreditação do Inmetro

#### 3.1. Contabilizar *features* customizadas

Além das técnicas de tokenização comumente utilizadas em pré-processamento de texto como TF-IDF e BOW, este trabalho pretende demonstrar que a adição de *features* customizadas na etapa de pré-processamento otimiza o resultado final. As *features* customizadas são parâmetros quantitativos assim como as outras *features* que compõem o TF-IDF. Elas foram identificadas através de análise prévia e conhecimento técnico sobre as respostas esperadas dos avaliadores, são elementos que o corpo técnico do Inmetro considera de grande importância estarem presentes para que as respostas estejam adequadas. As *features* customizadas podem ser visualizadas na Tabela 1.

Tabela 1. *Features* Customizadas

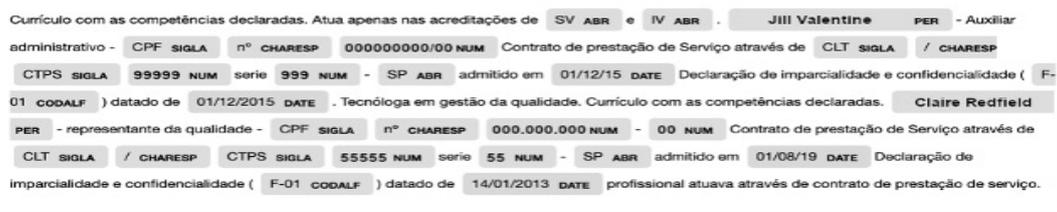
<i>Feature</i>	Tipo de informação	Exemplos
abreviaco0	Abreviações	MQ (manual da qualidade); OI (organismo de inspeção); EO (evidência objetiva)
siglas1	Siglas	CREA, CFT e CTPS
datas2	Datas	15/03/2020; período de 08/2018 a 02/2019; ciclo de 2017 a 2020; 22 e 23/02/2016
codalfas3	Códigos alfanúmericos	Certificado de Proficiência - Regloscópio - <b>RE-02-17-01</b> ; Número de inscrição <b>RI3.EAN.0.8.1P</b>
medidas4	Medidas	0,1 mm; 94 dB(A); 08hs; 1 kPa
nomes5	Nomes próprios e de empresas	João da Silva; Inspeciona bem Ltda
endinternet6	Endereços de internet	https://webapp1.otmws.com.br/SIVPP2/SivppEmpresa/Login.aspx sftp://35.170.103.199
charsespeciais7	Caracteres especiais	Nº, R\$, =, /
numeros8	Números	Nº 754935, R\$ 10000,00; Média de inspeções mês = 30,75 Média de inspeções / dia = 1,39; item 4.8.1: 0,1 mm; 94 dB(A); 08hs; 1 kPa

São 9 tipos diferentes de *features* customizadas: abreviações, siglas, datas, códigos alfanuméricos, unidades de medida, nomes próprios e de empresas, endereços de internet, caracteres especiais e números. A presença dessas *features* nos dados de treinamento indicará o peso de cada uma delas no momento de avaliar uma nova resposta.

Para a contabilização das *features* customizadas do texto foi utilizada uma biblioteca livre e de código aberto para processamento de linguagem natural em Python, chamada *spaCy*. A biblioteca possui um módulo para reconhecimento de entidades nomeadas (NER) em português, esse módulo, que já reconhece algumas entidades, pode ser incrementado através de treinamento ou pode-se eliminar todo seu conhecimento prévio de reconhecimento de entidades e ser treinado desde o início. Como as características buscadas no texto são bem específicas para ele, o módulo foi reiniciado e treinado desde o início. Essa etapa de treinamento do módulo NER precisou ser feita apenas uma vez, então não faz parte do fluxo de criação e validação dos classificadores, é um pré-requisito para que o fluxo seja realizado.

[Partalidou et al. 2019] explica o funcionamento do *spaCy* ao realizar NER. Cada parte do texto que pode representar um token ou uma classe do NER é representada por um ID. Prefixos, sufixos e forma são utilizadas para a extração de valores *hash* que indicarão as semelhanças entre as palavras. Os valores então são aplicados a uma CNN e são fundidos ao contexto, gerando uma matriz de vetores. Por fim, é utilizada uma função *Softmax* que leva em consideração classe gramatical e morfologia.

Como fonte de treinamento para o *spaCy* foram escolhidas 68 respostas dos relatórios existentes que possuíam as diferentes *features* customizadas desejadas, sendo que para cada uma das respostas foi necessário informar para o *spaCy* de qual posição até qual posição daquele texto o dado deveria ser considerado uma abreviação, um código alfanumérico, uma data, um nome de pessoa ou de empresa, um endereço de internet, etc. O resultado para as 3 perguntas foi bastante satisfatório e um exemplo pode ser visto na Figura 2. Devido a regras de confidencialidade, os nomes, datas e identificação de



The image shows a snippet of a document with several fields highlighted by gray boxes representing NER tags. The text includes: 'Currículo com as competências declaradas. Atua apenas nas creditações de SV ABR e IV ABR . JIII Valentine PER - Auxiliar administrativo - CPF SIGLA nº CHARESP 00000000/00 NUM Contrato de prestação de Serviço através de CLT SIGLA / CHARESP CTPS SIGLA 99999 NUM serie 999 NUM - SP ABR admitido em 01/12/15 DATE Declaração de imparcialidade e confidencialidade ( F-01 CODALF ) datado de 01/12/2015 DATE . Tecnóloga em gestão da qualidade. Currículo com as competências declaradas. Claire Redfield PER - representante da qualidade - CPF SIGLA nº CHARESP 000.000.000 NUM - 00 NUM Contrato de prestação de Serviço através de CLT SIGLA / CHARESP CTPS SIGLA 55555 NUM serie 55 NUM - SP ABR admitido em 01/08/19 DATE Declaração de imparcialidade e confidencialidade ( F-01 CODALF ) datado de 14/01/2013 DATE profissional atuava através de contrato de prestação de serviço.'

documentos foram alterados para informações fictícias.

**Figura 2. Exemplo do resultado da aplicação de um modelo NER a uma resposta dada à pergunta 6.1.2\***

### 3.2. Retirar *features* customizadas do texto

Todas as *features* customizadas contabilizadas na etapa anterior são retiradas do texto antes dos próximos passos de pré-processamento, justamente para que não sejam contabilizadas tanto como *feature* customizada quanto como *feature* gerada pela tokenização em 1-gram e 2-gram.

### 3.3. Outras etapas de pré-processamento de texto

Em seguida é aplicado o *lowercase* em todas as palavras do texto, ou seja, todas as letras ficam minúsculas. Todas as palavras consideradas *stopwords* da língua portuguesa são

retiradas do texto. Todos os números, sinais de pontuação, caracteres especiais e espaços em branco são retirados do texto.

Após todos esses passos são gerados os tokens nos formatos 1-gram e 2-gram e contabilizados, então é nesse momento que a contabilização das *features* customizadas realizada na primeira etapa é adicionada à contabilização dos tokens 1-gram e 2-gram.

Por fim, para finalizar o pré-processamento, é calculado o valor de TF-IDF de cada um dos tokens obtidos.

### 3.4. Obtenção dos modelos de classificação de texto

Para obtenção dos modelos de classificação de texto, foram aplicados os algoritmos de classificação *Naive Bayes* (NB), *Support Vector Machine* (SVM), a Regressão Logística com regularização Lasso (LR) e a Floresta Aleatória (RF), por se tratarem de algoritmos que possuem bom desempenho em altas dimensões e em conjunto de dados pequeno [Orozco *et al.*, 2018]. Além disso, algoritmos de *deep learning* foram desconsiderados devido à quantidade de relatórios existentes e rotulados não ser suficiente para um bom desempenho. [Wei *et al.* 2019] é um exemplo que demonstra que algoritmos tradicionais possuem desempenho melhor do que o *deep learning* no caso de *datasets* pequenos.

Os testes foram realizados por validação cruzada e estratificada *5-fold*, ou seja, o *dataset* foi quebrado em 5 *subsets*, e a cada iteração um *subset* era utilizado como dados de teste e os outros 4 como dados de treinamento. Por fim, foram geradas as curvas ROC e calculadas suas áreas. A curva ROC foi utilizada para avaliar o desempenho dos algoritmos, pois é adequada ao caso de classificação com custos diferentes para os erros [Fawcett, 2006].

### 3.5 Classificação dos elementos dos relatórios

Como mencionado anteriormente, cada relatório de acreditação do Inmetro possui 419 perguntas. As respostas dessas perguntas devem ser avaliadas para identificar estão adequadas ou inadequadas. Através do modelo de classificação obtido no passo anterior, realizar-se-á a análise dessas repostas automaticamente. Um modelo será construído para cada resposta. Tendo em vista que cada resposta possui informações diferentes.

## 4. Experimentos

Para a realização do estudo foram utilizados 586 relatórios do ano de 2019 e selecionou-se 3 perguntas. A Tabela 2 mostra as perguntas selecionadas para a criação dos classificadores. Elas foram selecionadas devido à existência de dados previamente rotulados em adequados e inadequados de forma manual. O total de dados nem sempre é 586 devido a alguns relatórios não terem algumas respostas preenchidas.

**Tabela 2. Perguntas selecionadas para o estudo**

Código	Pergunta	Qtd de respostas adequadas	Qtd de respostas inadequadas
6.1.2*	O OI emprega ou tem contrato com número suficiente de pessoas com as competências exigidas incluindo onde necessário habilidade para fazer julgamentos profissionais a fim de	410	170
6.1.9*	Cada inspetor é observado no local de trabalho ou caso negativo há evidência fundamentada pela combinação de informações de que o inspetor continua atuando com competência? (verificar as informações previstas na NIT-Diois-008).	319	265
6.2.4*	A definição dos equipamentos inclui a exatidão requerida e o intervalo de medição quando necessário?	275	309

## 4.1. Respostas adequadas e inadequadas

A Tabela 3 exibe alguns exemplos de respostas adequadas e inadequadas para as 3 perguntas selecionadas. Devido a regras de confidencialidade, os nomes, datas e identificação de documentos foram alterados para informações fictícias.

**Tabela 3. Perguntas e Respostas Adequadas e Inadequadas**

Código	Exemplo de resposta adequada	Exemplo de resposta inadequada
6.1.2*	Quadro técnico relacionado na aba (4) quadro Pessoal. RT Popeye (escopos autorizados PP + Descontaminação / SV / IV) - Carteira de trabalho assinada em 99/99/9999 - Certidão CREA n. 00000 - validade 11/11/1111 RT eventual Chapolin Colorado (escopos autorizados PP / SV / IV) - Contrato de Prestação de Serviços datado de 22/22/2222 (registrado no 4o Ofício RTD sob n. 11111 - Livro B em 55/55/5555) - Certidão CREA n. 22222 - validade 44/44/4444 Inspetor John Doe (escopos autorizados SV / IV) - Carteira de trabalho assinada em 88/88/8888 - Certidão CFT n. 7777777/7777 - validade 99/99/9999 Inspetor Jack Bauer (escopos autorizados SV / IV) - Carteira de trabalho assinada em 00/00/0000 - Certidão CFT n. 9999999/9999 - validade 66/66/6666 Inspetor Robocop (escopos autorizados SV / PP / IV) - Carteira de trabalho assinada em 88/88/8888 - sem comprovação de regularidade de registro junto ao conselho de classe - ver NC Inspetor Luke Skywalker (escopos autorizados PP / SV / IV) - Carteira de trabalho assinada em 22/22/2222 - Certidão CFT n. 0000000/0000 - validade 88/88/8888 Operador na Descontaminação - Chewbacca - Carteira de trabalho assinada em 55/55/5555 (NC 11/11 PP - Evidenciado que o OIA não apresentou comprovação de regularidade do inspetor Han Solo junto ao Conselho de Classe.)	Evidenciado vínculo empregatício de todos os colaboradores como por exemplo: 1) Spock - inspetor - vínculo CLT com o organismo desde 00/00/0000 evidenciado através de sua CTPS número 00000 - série 1111HH.
6.1.9*	Evidenciado no PA RH 01_05 e item 6.1.9 do MSC SQQ 01_11 que o RQ RT ou auditor interno na data da auditoria devem observar situações relativas aos riscos e a complexidade das inspeções resultados de monitoramentos das atividades anteriores desenvolvimentos técnicos processuais ou legais relevantes para determinar a realização de uma ou mais amostras para avaliação. Evidenciado Planejamento de avaliação de desempenho para as equipes técnica e administrativa devidamente registrado no FR RH 06 - Programa de Avaliação de Desempenho. E.O.: Mickey Mouse - Inspetor - avaliação realizada em 00/00/0000 por Pato Donald RT na amostra placa TTT 2222 - Portaria 32/04.(verificado vídeo de monitoramento)	Todos os inspetores são avaliados no local de trabalho conforme o programa de avaliação de desempenho e registros anotados no item anterior do RAO.
6.2.4*	Sistemática no item 7.3 do PA-08- rev.06 - IDENTIFICAÇÃO REGISTRO E CALIBRAÇÃO DE EQUIPAMENTOS DE INSPECÇÃO - EIMES. EO: Paquímetro - Patrimônio Nº 000-000 - Erro Máximo / Exatidão: ± 005 mm (Conforme a Tabela 2 da NBR NM-216:2000) - Faixa de medição de 0 à 150 mm.	Conforme relato do avaliador técnico o OI possui sistemática esta definida no FR EQP 02 - Plano de Calibração de Equipamentos de Medição conforme descrito no PA EQP 01_06.

Verifica-se que uma resposta adequada para a pergunta 6.1.2\* possui informações de diversas pessoas com as competências exigidas, forma e data de contratação, códigos de contratos, de carteiras de trabalho, e de certificados. Já o exemplo inadequado é um texto com poucas informações, indicando apenas um dos colaboradores do organismo.

Para a pergunta 6.1.9\*, verifica-se que uma resposta adequada possui informações sobre o trabalho do inspetor avaliado, como nome, data da avaliação, e dados da amostra. Além disso, foi dito que o trabalho do inspetor foi verificado através de vídeo de monitoramento. Já o exemplo inadequado apenas cita genericamente que os inspetores são avaliados conforme programa de avaliação de desempenho, não mostrando nenhuma informação aprofundada de que há evidências de atuação com competência.

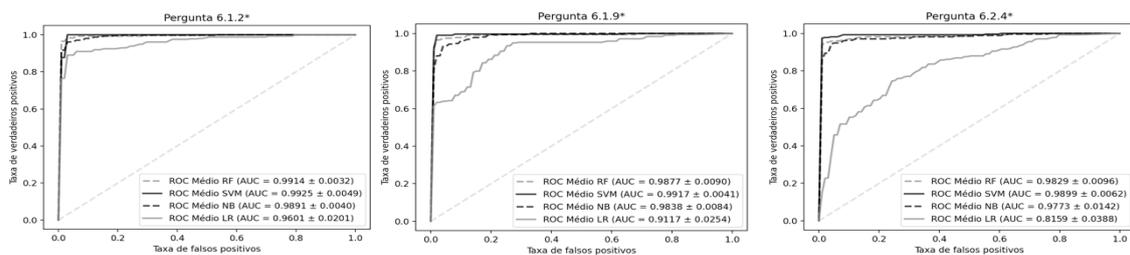
Por fim, para a pergunta 6.2.4\*, uma resposta adequada precisa possuir, além de uma resposta afirmativa ou negativa, alguns dados que corroborem o que está sendo afirmado. No exemplo adequado é fornecido dados de um equipamento, como número de patrimônio, intervalo de medição e erro máximo. Já a resposta inadequada apenas informa que o organismo possui sistemática definida no plano de calibração de equipamentos de medição, gerando dúvida em relação à checagem do avaliador.

## 4.2. Implementação dos Algoritmos

Para execução dos algoritmos, utilizou-se o programa Python e a biblioteca *scikit-learn*. Os hiperparâmetros dos algoritmos foram definidos da seguinte forma: a) RF – número de estimadores = 200, profundidade máxima = 3, número mínimo de divisões das amostras = 2 e semente para geração de números aleatórios = 0; b) SVM – kernel linear; c) NB – modelo multinomial; d) LR – penalização = “11”, C = 0,1, semente para geração de números aleatórios = 42. Os demais hiperparâmetros utilizaram os valores padrão da biblioteca.

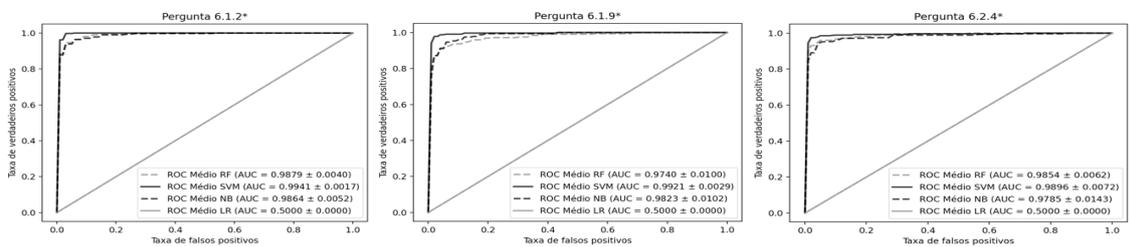
## 4.3. Desempenho dos classificadores com e sem *features* customizadas

A performance dos classificadores foi avaliada com a medição da área sob a curva ROC (AUC). A Figura 4 mostra as curvas ROC e os valores médios de AUC para os algoritmos RF, SVM, NB e LR considerando as *features* customizadas na análise das respostas para as perguntas 6.1.2\*, 6.1.9\* e 6.2.4\*.



**Figura 4. Curva ROC para as perguntas 6.1.2\*, 6.1.9\* e 6.2.4\* para os algoritmos RF, SVM, NB e LR considerando as *features* customizadas. À esquerda a pergunta 6.1.2\*, no centro a pergunta 6.1.9\* e à direita a pergunta 6.2.4\***

Por sua vez, a Figura 5 exibe as curvas ROC e os valores de AUC para os algoritmos RF, SVM, NB e LR sem as *features* customizadas na análise das respostas.



**Figura 5. Curva ROC para as perguntas 6.1.2\*, 6.1.9\* e 6.2.4\* para os algoritmos RF, SVM, NB e LR sem as *features* customizadas. À esquerda a pergunta 6.1.2\*, no centro a pergunta 6.1.9\* e à direita a pergunta 6.2.4\***

As curvas foram desenhadas considerando os valores médios das iterações da validação cruzada com 5 divisões de conjuntos. Os valores de AUC foram calculados através da média dos valores de AUC referentes aos 5 conjuntos de validação.

Através das Figuras 4 e 5, nota-se que o SVM obteve o melhor desempenho para as 3 perguntas considerando o valor médio de AUC em comparação aos demais algoritmos. Além disso, constata-se que as *features* customizadas tiveram impacto positivo no desempenho a depender do algoritmo e do tipo de pergunta. O algoritmo LR tem melhora no desempenho com a presença das *features* customizadas para todas as perguntas. Porém, no caso do SVM, apenas na pergunta 6.2.4\* ocorre melhoria devido à presença das *features* customizadas. Por sua vez, os algoritmos RF e NB melhoram seu desempenho com as *features* customizadas para as perguntas 6.1.2\* e 6.1.9\* apenas.

## 5. Conclusão

Neste trabalho foram apresentadas duas contribuições: um procedimento para análise automática de relatórios de acreditação utilizados pelo Inmetro; e aplicação de *features* customizadas para melhoria do desempenho dos algoritmos na obtenção de modelos capazes de analisar os relatórios. Considerando os algoritmos RF, SVM, NB e LR, o SVM obteve o melhor desempenho perante métrica AUC na análise das respostas das perguntas sob estudo. Além disso, a utilização de *features* customizadas teve impacto positivo na melhoria do desempenho a depender do algoritmo e do tipo de pergunta abordada.

Essas constatações dão indícios de que o algoritmo SVM seria o mais indicado para obtenção de modelos de classificação no caso de relatórios de acreditação do Inmetro. Ademais, há indícios de que para cada pergunta exista a necessidade de se desenvolver *features* customizadas diferentes.

Para trabalhos futuros, sugere-se: a investigação de outras *features* customizadas; modelagem para outras perguntas dos relatórios; e estabelecimento de um mecanismo que indique o motivo da reprovação da resposta de uma pergunta para auxiliar o usuário final.

## Referências

- ALETRAS, N.; TSARAPATSANIS, D.; PREOȚIUC-PIETRO, D.; LAMPOS, V. Predicting judicial decisions of the European court of human rights: A natural language processing perspective. **PeerJ Computer Science**, v. 2016, n. 10, p. 1–19, 2016.
- BAHGAT, E. M.; RADY, S.; GAD, W.; MOAWAD, I. F. Efficient email classification approach based on semantic methods. **Ain Shams Engineering Journal**, v. 9, n. 4, p. 3259–3269, 2018.
- CACCAMISI, A.; JØRGENSEN, L.; DALIANIS, H.; ROSEN LUND, M. Natural language processing and machine learning to enable automatic extraction and classification of patients' smoking status from electronic medical records. **Uppsala Journal of Medical Sciences**, v. 125, n. 4, p. 1–9, 2020.
- CARUANA, R.; KARAMPATZIAKIS, N.; YESSENALINA, A. An empirical evaluation of supervised learning in high dimensions. **Proceedings of the 25th International Conference on Machine Learning**, p. 96–103, 2008.
- INMETRO - Instituto Nacional de Metrologia, Qualidade e Tecnologia, **Acreditação**. Disponível em: <<https://www.gov.br/inmetro/pt-br/assuntos/acreditacao/cgcre/acreditacao>>. Acesso em: 15 abr. 2021.
- FAWCETT, T. An introduction to ROC analysis. **Pattern recognition letters**, v. 27, n. 8, p. 861-874, 2006.
- OROZCO, L. E. J., MANZANERA, O. M., NESTEROV, S. V., KAJANDER, S., KINUTI, J. The machine learning horizon in cardiac hybrid imaging. **European Journal of Hybrid Imaging**, v. 2, n. 1, p. 1-15, 2018.
- LEI, M.; GE, J.; LI, Z.; et al. Automatically classify Chinese judgment documents utilizing machine learning algorithms. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 10179 LNCS, n. March, p. 3–17, 2017.
- PALATUCCI, M.; MITCHELL, T. M. Classification in very high dimensional problems with handfuls of examples. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 4702 LNAI, p. 212–223, 2007.
- PARTALIDOU, E.; SPYROMITROS-XIOUFIS, E.; DOROPOULOS, S.; VOLOGIANNIDIS, S.; DIAMANTARAS, K. I. Design and implementation of an open source Greek pos tagger and entity recognizer using spaCy. **arXiv**, p. 337–341, 2019.
- UKEY, K. P.; ALVI, A. S. Text Classification Using Support Vector Machine with Mixture of Kernel. **International Journal of Engineering Research & Technology (IJERT)**, v. 1, n. 3, p. 55–58, 2012.
- WEI, F.; QIN, H.; YE, S.; ZHAO, H. Empirical Study of deep learning for text classification in legal document review. **arXiv**, 2019.