

**Apreciação de Procedimentos para Apuração de Descumprimento de
Obrigações – PADOs sobre Serviços de Telecomunicações**

Utilizando o Método da Árvore de Decisão para Analisar o Comportamento dos
Infratores

Trabalho de Conclusão de Curso apresentado
como parte dos requisitos para obtenção do grau
de Especialista em Análise de Dados em
Políticas Públicas.

Aluna: Herika Kawata

Orientador: Professor Doutor Eduardo Monteiro
de Castro Gomes

Brasília – DF
setembro/2019

Apreciação de Procedimentos para Apuração de Descumprimento de Obrigações – PADOs sobre Serviços de Telecomunicações

Autora: Herika Kawata

Instituição: Agência Nacional de
Telecomunicações – Anatel

Palavras chave: infrações administrativas, árvore de decisão, telecomunicações

Resumo analítico

Os Procedimentos para Apuração de Descumprimento de Obrigações – PADOs são processos sancionatórios instaurados para apurar infrações de competência da Agência Nacional de Telecomunicações – Anatel.

A avaliação do comportamento dos infratores frente a multas aplicadas viabiliza o controle sobre a atividade administrativo-sancionadora, permitindo ajustes na metodologia para melhorar sua eficácia frente a sociedade.

Este trabalho objetiva compreender o processo decisório das prestadoras infratoras em relação aos PADOs a fim de inibir o cometimento de ilícitos administrativos.

O método escolhido é a árvore de decisão, a qual permitirá a visualização de possíveis resultados de uma série de escolhas relacionadas. As árvores de decisão (decision trees) trabalham com problemas complexos mostrando graficamente os caminhos da decisão tomada, tornando mais fácil a análise de eventos futuros trabalhadas nos “ramos” das árvores.

Espera-se que os resultados desta análise possam contribuir para melhorar a efetividade do sancionamento das empresas, atuando de acordo com as boas práticas da regulação responsiva.

Por fim, pretende-se fornecer informações relevantes para aprimorar a atuação regulatória da Anatel em relação à qualidade dos serviços prestados à sociedade.

Introdução

Este artigo tem por objetivo abordar a aplicação de sanções na Anatel, em especial analisando o comportamento dos infratores frente a atuação da Anatel em relação aos processos sancionatórios com a penalidade de multa aplicada.

Compreendendo melhor o processo decisório das prestadoras infratoras, espera-se que os resultados contribuam para avaliar a efetividade dessas sanções e indicar possíveis adequações do seu ato normativo.

A complexidade das políticas que envolvam várias decisões exige o uso de ferramentas analíticas que possam dissecar o problema dando ao gerente público uma perspectiva mais realista das decisões a serem tomadas (CARVALHO, 2005).

Uma das principais competências da Anatel é arbitrar conflitos entre o poder concedente, os concessionários e os usuários, aplicando sanções administrativas quando necessário.

Das sete sanções previstas no Regulamento de Aplicação de Sanções Administrativas da Anatel (Resolução nº 589, de 7 de maio de 2012), a aplicação de sanção pecuniária é sem dúvida a mais utilizada no âmbito da Agência.

Infelizmente, grande parte dessas multas não são pagas pelos infratores, cabendo a interposição de recurso administrativo ou pedido de reconsideração, bem como seu encaminhamento à esfera judicial.

Diante desta conjuntura, o ato normativo citado acima inseriu um dispositivo que permite um desconto no pagamento da multa aplicada, desde que o infrator renuncie expressamente de seu direito de recorrer, *in verbis*:

“Art. 33. O pagamento da multa deve ser efetuado no prazo de 30 (trinta) dias, contado a partir do recebimento da intimação da decisão de aplicação de sanção.

(...)

§ 5º O infrator que renunciar expressamente ao direito de recorrer da decisão de primeira instância, fará jus a um fator de redução de 25% (vinte e cinco por cento) no valor da multa aplicada, caso faça o recolhimento no prazo para pagamento definido no caput.”

Este trabalho analisará os processos com multas aplicadas pela Anatel, avaliando quais as características dos processos bem como os critérios utilizados pelos infratores para a utilização do desconto no pagamento, renunciando ao seu direito de defesa.

Por fim, pretende-se fornecer informações relevantes para aprimorar a conformidade regulatória da Anatel, adequando sua fiscalização responsiva e consequentemente impactando na qualidade dos serviços prestados à sociedade.

Metodologia escolhida

Neste estudo será utilizada uma das ferramentas de aprendizado supervisionado, o qual envolve a construção de um modelo estatístico para prever ou estimar uma resposta de acordo com uma ou mais informações de entrada.

A variável resposta ou variável dependente é a renúncia ou não dos processos com multas aplicadas, conforme explicitado anteriormente, pois o objetivo é estudar a previsibilidade dessa situação ocorrer baseado no histórico da Agência.

O algoritmo escolhido é a árvore de decisão (*decision tree*), que mostra a estrutura lógica de um problema, podendo ser utilizada apenas para representar uma decisão ou pode auxiliar na investigação da mesma.

Por se tratar de uma análise inédita sobre o tema e desta forma tornar necessário eventualmente um estudo posterior mais aprofundado, optou-se por um modelo mais intuitivo, que prioriza a classificação das observações. As árvores de decisão identificam a variável mais representativa e os valores que retornam os conjuntos de população mais homogêneos.

Esta metodologia foi escolhida devido às seguintes vantagens:

- É considerado um modelo não-paramétrico, isto é, não assume nenhuma distribuição para os dados, fornecendo uma cobertura exaustiva do espaço de objetos;
- Possui relativa robustez a valores atípicos e ruídos, uma vez que a construção da árvore inclui o processo de seleção de atributos relevantes;
- Por fim, possui boa interpretabilidade, sendo mais fácil de explicar e compreender em relação a outros modelos similares.

Os termos mais comuns utilizados para árvores de decisão e que serão usados para a análise dos dados deste trabalho são os seguintes:

- Nó Raiz: população ou amostra estudada;
- Divisão: processo de dividir um nó em dois ou mais sub-nós;
- Nó de decisão: divisão de um sub-nó em sub-nós adicionais;
- Folha ou nó de término: são os nós não divididos;
- Poda: processo de remover sub-nós de um nó de decisão;

- Ramificação ou sub-árvore: sub-seção da árvore inteira;
- Nó pai e nó filho: um nó que é dividido em sub-nós é chamado de nó pai. Os sub-nós são os nós filhos do nó pai.

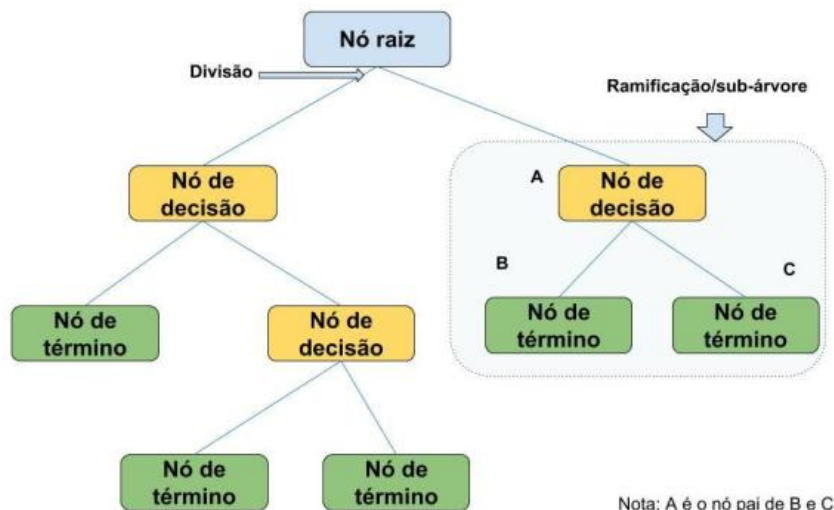


Figura 1 – Representação gráfica das terminologias em uma árvore de decisão.

Para a criação do algoritmo, foram analisados todos os PADOs encerrados com a penalidade multa aplicada instaurados a partir de 2013, com ênfase nos processos dos grandes grupos econômicos que atuam no mercado de telecomunicações.

Ademais, optou-se por avaliar apenas quatro tipos de serviços de telecomunicação, quais sejam: telefonia móvel (Serviço Móvel Pessoal – SMP), telefonia fixa (Serviço Telefônico Fixo Comutado – STFC), internet banda larga (Serviço de Comunicação Multimídia – SCM) e tv por assinatura (Serviço de Acesso Condicionado – SeAc).

Este corte temporal em 2013 foi necessário em virtude de uma particularidade em relação ao contexto da Agência naquele momento, pois anterior a este período não havia uma comunicação automática entre o sistema que administrava os processos sancionatórios e o que gerenciava a arrecadação da Agência, incluindo as multas aplicadas. Anteriormente, as informações sobre o andamento dos processos eram inseridas manualmente no sistema de arrecadação, sem garantia de confiabilidade dos dados. Foi somente a partir de meados de 2012 que os números de processo passaram a ser inseridos no sistema de arrecadação da Anatel de forma automatizada, possibilitando a rastreabilidade das informações.

Além disso, o mais recente ato normativo que regulamenta a aplicação de sanções administrativas da Anatel foi publicado em maio de 2012, sendo mais apropriado

considerar os processos instaurados a partir de 2013 para a análise do desconto no pagamento da multa, conforme explicitado acima.

Por fim, é importante ressaltar que a partir de 2016 a Anatel adotou o sistema de processo eletrônico, facilitando a extração de dados sobre os processos estudados e permitindo que este tipo de estudo fosse realizado.

Desenvolvimento

O Estado Moderno trouxe uma nova forma de atuação do Estado no plano econômico, diminuindo sua atuação empreendedora e transferindo sua responsabilidade principal para o campo da regulação e fiscalização dos serviços delegados.

Neste novo paradigma surgiram as agências reguladoras, que desempenham funções executivo-administrativas, normativas e decisórias, dentro de um espaço de competências deferido por lei.

Desta forma, a Agência Nacional de Telecomunicações – Anatel é criada em 1997 com a competência de organizar a exploração dos serviços de telecomunicações, em especial quanto aos aspectos de regulamentação e de seu acompanhamento, outorga de concessão e permissão, expedição de autorização, uso dos recursos de órbita e de radiofrequências, fiscalização e aplicação de sanções.

Conforme Lodge e Wegrich (2012), a regulação é uma atividade intencional que procura alterar o comportamento de outra parte. Ela inclui três componentes: normatização (para fornecer direcionamento geral), ferramentas de modificação de comportamento (para o atingimento de padrões estabelecidos) e coleta de informações (para permitir o conhecimento sobre o que está acontecendo). Os autores reforçam ainda que os regimes regulatórios só podem alcançar seus efeitos se todos os três componentes operam de maneira funcional.

Neste contexto, as agências reguladoras no Brasil e, particularmente a Anatel, passam por um momento de aprimoramento e modernização de seu processo de regulamentação, adequando-se às novas práticas regulatórias para maior previsibilidade, transparência e estabilidade de seu processo. Conceitos como responsividade, atuação com base em evidência, avaliação de risco e dos resultados têm sido incorporados nas Agências visando conformidade regulatória e melhora da qualidade na prestação dos serviços delegados. Isto sem dúvida têm influenciado a mentalidade tradicional de sancionamento, abrindo espaço para a fiscalização responsiva.

O princípio de fiscalização responsiva foi introduzido originalmente por Ayres e Braithwaite (1992) e estabelece que as estratégias para obtenção de conformidade regulatória devem ser moduladas a partir do comportamento e histórico das empresas fiscalizadas, em ações escalonadas de intervenções persuasivas e dissuasivas de regulação.

Uma das ideias centrais da teoria de fiscalização responsiva costuma ser representada esquematicamente pelos autores e por diversas instituições que a aplicam por meio da “pirâmide de conformidade regulatória” expressa visualmente na Figura 1 a seguir.

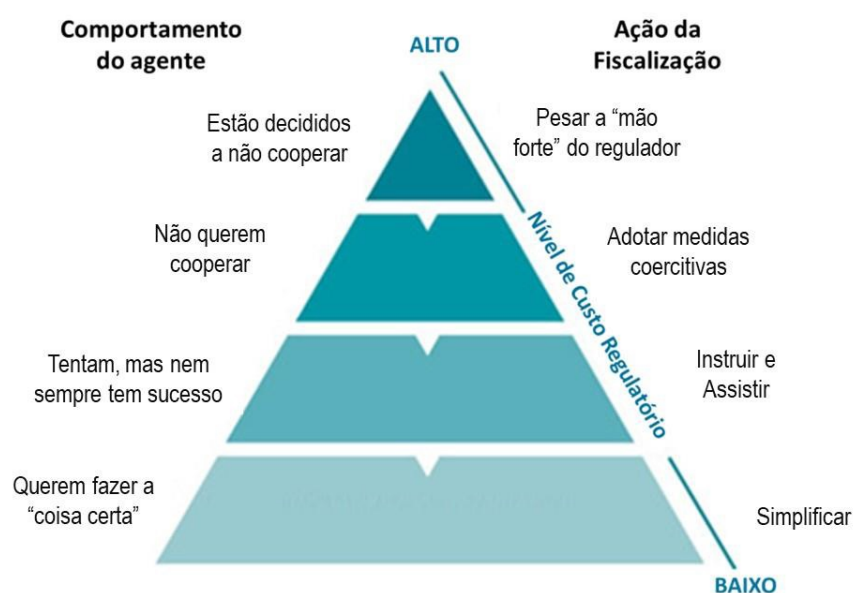


Figura 2 – Pirâmide de conformidade regulatória, esquematização da teoria de regulação responsiva.

Este trabalho pretende subsidiar a Agência com informações relevantes para serem utilizadas na fiscalização responsiva, indicando possíveis caminhos de intervenção de acordo com o nível de custo regulatório.

Análise dos dados

Conforme explicitado anteriormente, somente os processos instaurados a partir de 2013 dos serviços de telecomunicação citados foram considerados nesta análise. Os dados abaixo foram extraídos da base de dados da Anatel em julho de 2019. Desta forma, foram encontrados 4.650 PADOs instaurados neste período. Deste montante, 2.486 foram registrados no sistema de arrecadação da Anatel, ou seja, estes processos

sancionatórios tiveram multa aplicada. Para a análise deste artigo, serão considerados apenas os processos encerrados até julho de 2019, com um total de 2.010 processos.

Inicialmente, 8 variáveis foram extraídas da base de dados da Anatel para análise. As mesmas foram escolhidas devido a sua importância nos processos sancionatórios da Agência e pela possível relevância para este estudo. É necessário ressaltar também que por se tratar de um modelo preditivo, apenas as variáveis que estão à disposição no processo antes da decisão da renúncia foram levadas em consideração. São as seguintes:

- Número de processo: variável selecionada apenas para fins de identificação e distinção de cada registro;
- Superintendência responsável: para este estudo serão tratadas como “Superintendência 1” e “Superintendência 2”;
- Ano de instauração do processo: considerou-se a data de assinatura do documento de instauração do processo;
- Tipologia do processo: é a classificação dos processos instaurados via sistema de processo eletrônico (SEI). Para este estudo foram selecionadas as 21 tipologias mais relevantes para os processos sancionatórios, as quais representam 85% do total de registros estudados, as demais foram marcadas como “Outros”;
- Grupo econômico da empresa infratora: para fins de análise da relevância desta variável, foram considerados os 7 grupos econômicos que mais apareceram nesta base de dados, os quais representam 9% do total de infratores estudados, os demais foram marcados como “Outros”;
- Serviço de telecomunicações: conforme dito anteriormente foram considerados apenas os seguintes serviços: telefonia móvel (Serviço Móvel Pessoal – SMP), telefonia fixa (Serviço Telefônico Fixo Comutado – STFC), internet banda larga (Serviço de Comunicação Multimídia – SCM) e tv por assinatura (Serviço de Acesso Condicionado – SeAc);
- Renúncia de defesa (desconto no pagamento): para fins deste estudo considera-se “Não” nos casos em que não houve renúncia ao direito de defesa e “Sim” como direito de defesa renunciado e consequente pagamento da multa com desconto;
- Valor da multa aplicada: valor da multa principal, isto é, sem atualização monetária e/ou juros moratórios. No caso de pagamento da multa com desconto foi considerado o valor original – sem a aplicação do desconto – a fim de analisar o impacto do valor monetário na decisão dos infratores;

É importante ressaltar que dados sobre a entidade e informações patrimoniais não foram considerados por se tratarem de dados sigilosos. Ademais, a informação sobre o grupo econômico no qual está inserida a entidade pode ser considerada suficiente para os propósitos desta análise.

A variável resposta deste estudo é a “Renúncia”, a qual indica se o processo teve ou não renúncia do direito de defesa. Neste caso, a multa deve ser quitada em 30 dias com desconto de 25% do valor original.

Em relação aos 2.010 processos analisados, em 22% foram feitos pedidos de renúncia, conforme se observa no gráfico abaixo:

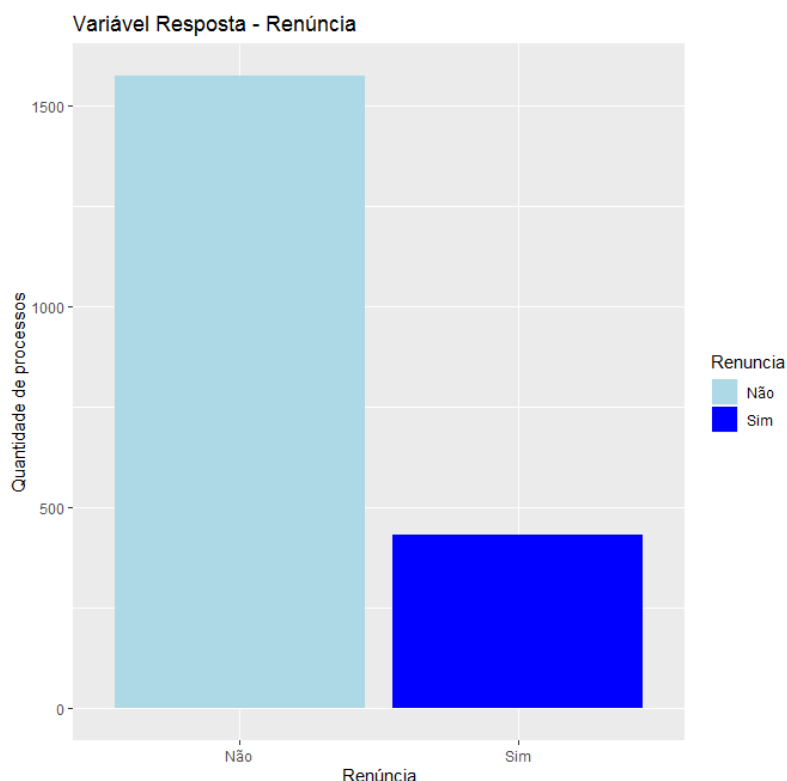


Gráfico 1: Distribuição dos processos de acordo com a variável “Renúncia”

Primeiramente, para as variáveis categóricas foi realizado um teste de aderência com o qui-quadrado (χ^2) de Pearson juntamente com a variável dependente “Renúncia”, a fim de avaliar a independência de cada uma neste contexto, conforme compilado abaixo:

Variável	χ^2	Graus de Liberdade	Probabilidade de significância
Superintendência	10.807	df = 1	p-value = 0.001011
Ano de instauração	94.068	df = 6	p-value = 0.152
Tipologia do processo	32.81	df = 21	p-value = 0.04837
Serviço de telecomunicação	42.34	df = 3	p-value = 3.398e-09
Grupo econômico	76.95	df = 7	p-value = 5.759e-14

Tabela 1: Teste qui-quadrado (χ^2) de Pearson com correção de continuidade de Yates

Observa-se que há rejeição à hipótese nula de que não há associação entre as variáveis, isto é, não há dependência entre as variáveis estudadas em relação a variável dependente.

Há duas Superintendências na Anatel responsáveis pelos processos de sanção. Para fins de estudo, serão denominadas como “Superintendência 1” e “Superintendência 2”. A distribuição dos processos neste contexto é dada da seguinte forma:

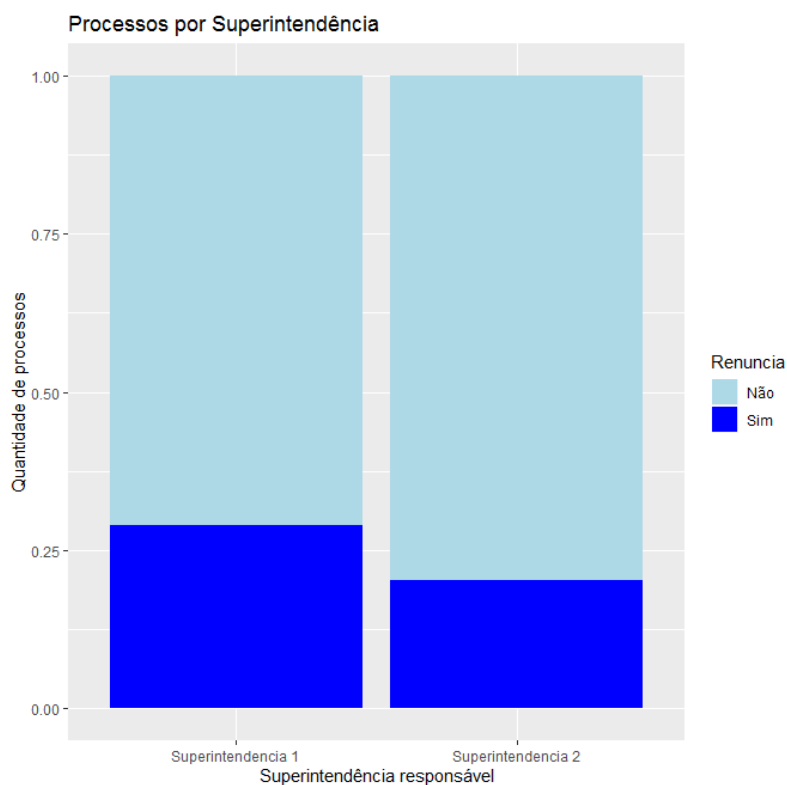


Gráfico 2: Processos por Superintendência de acordo com a variável “Renúncia”

Em relação ao ano de instauração dos processos, o gráfico abaixo indica que não houve alterações significativas de um período a outro para pedidos de renúncia, oscilando entre 20% a 25% do total de processos instaurados no ano.

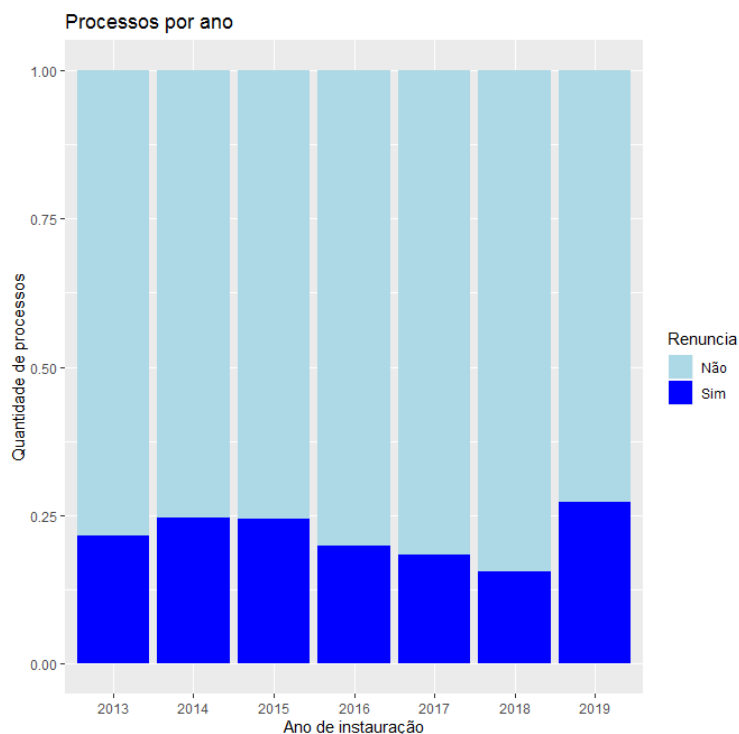


Gráfico 3: Processos por ano de instauração de acordo com a variável “Renúncia”

A distribuição dos processos em relação ao tipo de serviço de telecomunicação estudado indica que há uma quantidade menor de renúncias em processos de infrações de internet de banda larga (Serviço de Comunicação Multimídia – SCM), conforme se observa abaixo, em relação aos demais serviços multados:

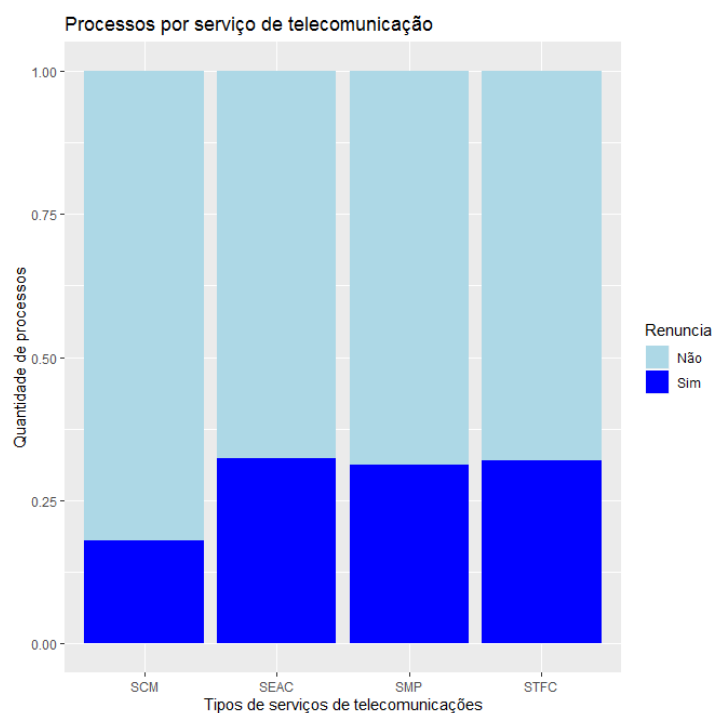


Gráfico 4: Processos por serviço de telecomunicações de acordo com a variável “Renúncia”

Em relação aos grupos econômicos foram selecionados os 7 mais recorrentes nesta análise, os quais representam apenas 9% do total de entidades infratoras. O restante dos processos (cerca de 1.720 processos) é formado em sua maioria por entidades que não pertencem a nenhum grupo, ou seja, pequenas empresas ou mesmo pessoas físicas que praticaram alguma infração.

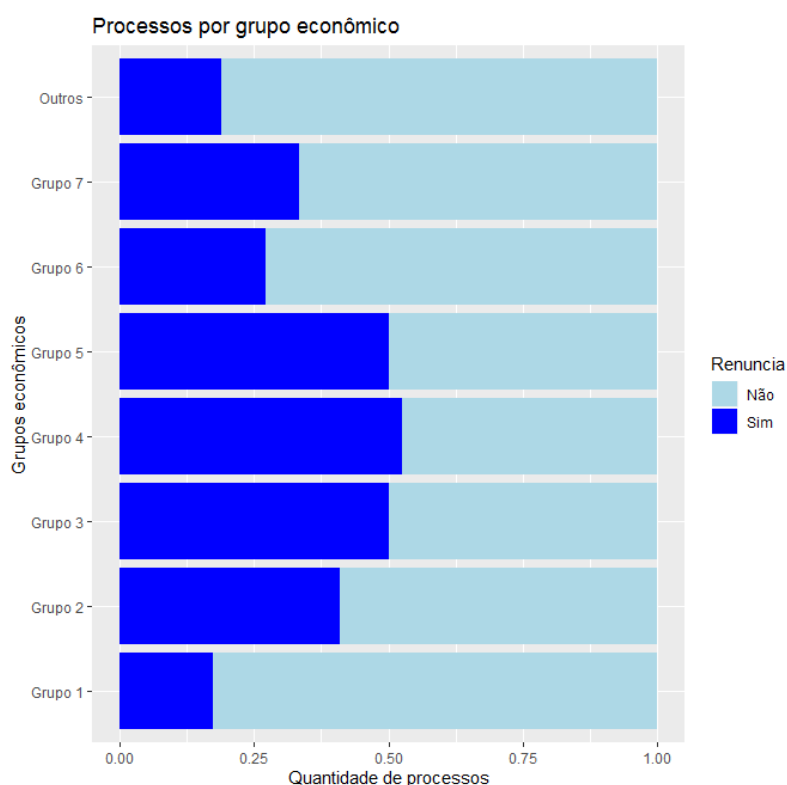


Gráfico 5: Processos por grupo econômico de acordo com a variável “Renúncia”

Em relação ao atributo “Tipologia dos processos”, percebe-se que dos 21 tipos de processos analisados, a maior parte das renúncias ocorreram nas tipologias: “Universalização”, “Gestão da qualidade – STFC”, “Direitos do consumidor – STFC” e “Atendimento – SCM”. No entanto, é importante ressaltar que a maior quantidade de processos está relacionada a irregularidades técnicas, isto é, infrações de não conformidade de instalações, equipamentos, aplicativos, sistemas, recursos e facilidades tecnológicas, documentos, dados e demais informações de natureza técnica regulamentados pela Anatel, os quais pertencem às tipologias: “Não outorgado – Serviço”, “Certificação e não outorgado – Radiofrequência”, “Licenciamento de Estação”, “Obrigações legais e contratuais”, dentre outros.

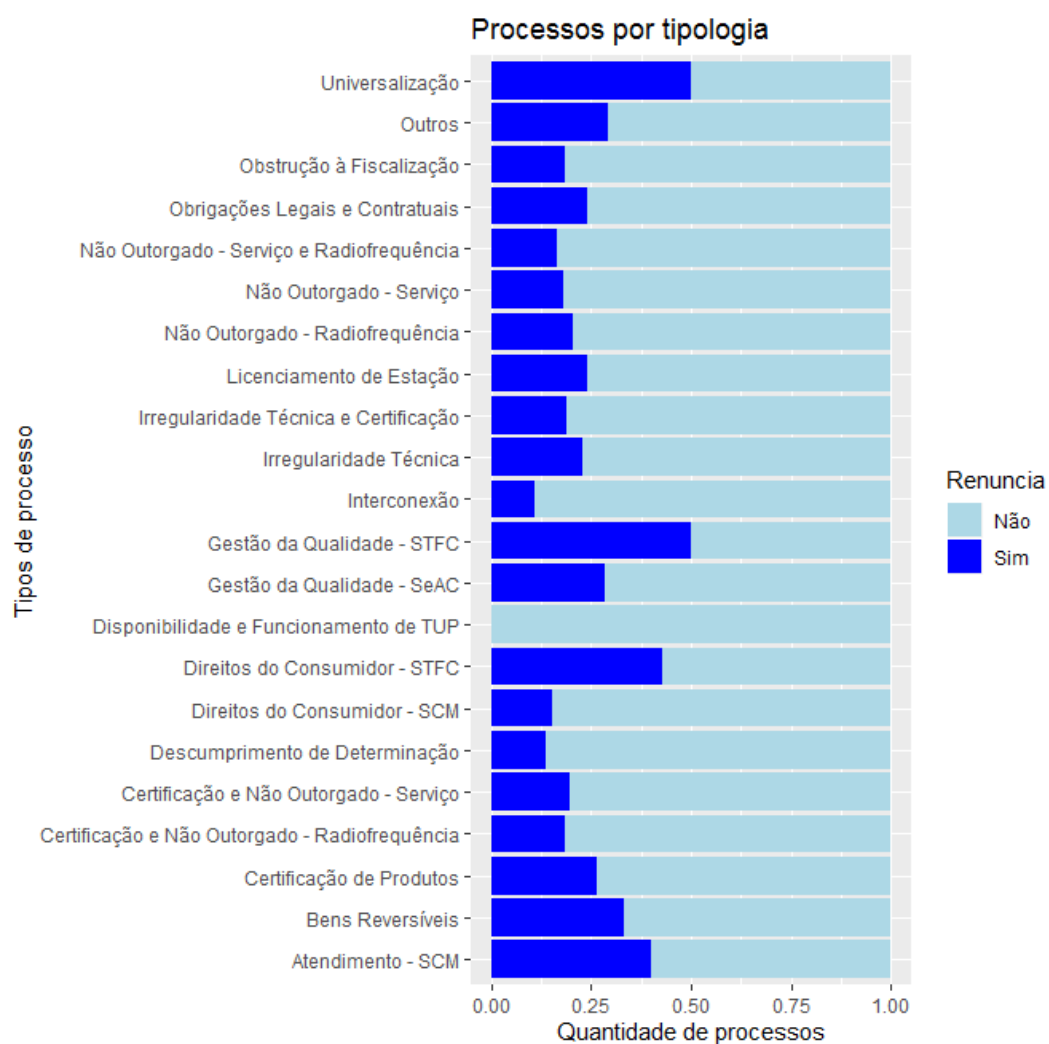


Gráfico 6: Processos por tipo de processo de acordo com a variável “Renúncia”

Em relação aos valores de multa, percebe-se que há uma grande amplitude dos valores mínimos e máximos, gerando outliers conforme se observa no gráfico abaixo:

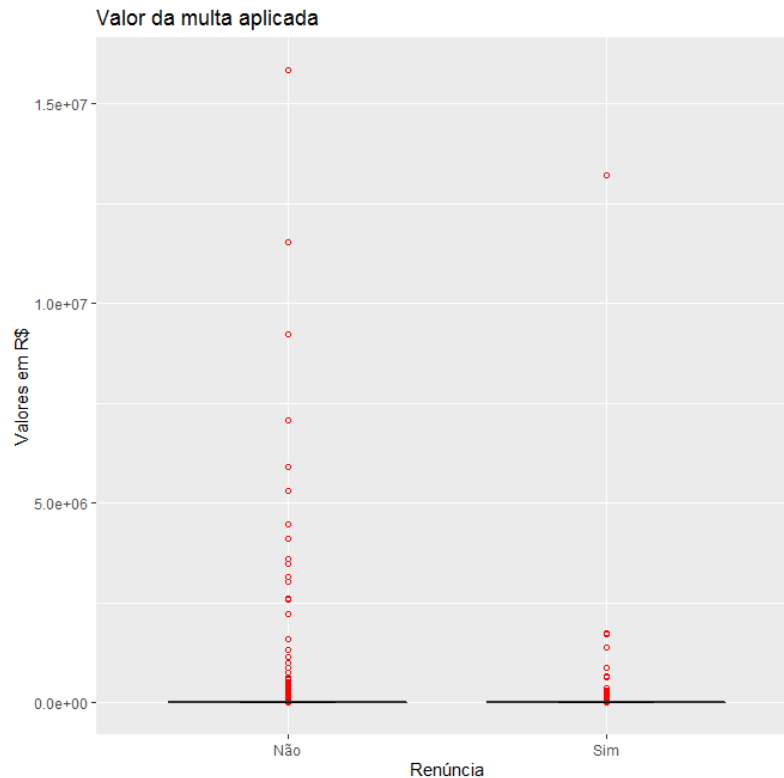


Gráfico 7: Boxplot dos valores de multa de acordo com a variável “Renúncia”

Conforme citado anteriormente, como o método da árvore de decisão é considerado robusto o suficiente para minimizar o efeito da presença de ruídos nos dados, os outliers não foram tratados para esta análise.

Construção do modelo

A base de dados estudada possui uma quantidade suficientemente pequena de observações com renúncia de defesa (22% dos seus processos). Assim, de forma a estressar o modelo para que atenda a todas as possibilidades, foram criadas 2 amostras de treinamento e teste: uma com dados balanceados (utilizando o balanceamento do tipo *oversampling*) e outra amostra sem balanceamento. O resultado dos dados balanceados se mostrou muito similar ao dos dados não balanceados quando a base de treino é utilizada, porém na base de teste se mostrou ineficaz na predição dos resultados, com taxa de acerto em torno de 35%. Desta forma, optou-se por demonstrar apenas os resultados da base não balanceada. A codificação de todas as etapas, inclusive as avaliações de todos os modelos nas duas amostras estão compiladas no Anexo II deste trabalho.

Para a construção da árvore foram utilizadas as seguintes variáveis: “Superintendência”, “Ano de instauração”, “Grupo econômico”, “Serviço de telecomunicação”, “Tipologia do processo” e “Valor da multa”. A variável “Número de processo” foi utilizada apenas para fins de identificação e distinção dos registros extraídos da base de dados.

O nó raiz, isto é, a amostra estudada, são todos os processos encerrados até julho de 2019 que foram instaurados a partir de 2013 com a penalidade de multa aplicada. Assim, serão considerados os registros dos 2.010 processos encontrados neste contexto. A divisão da árvore será realizada pela variável resposta “Renúncia”, buscando padrões para a criação de nós de decisão ou para sub-nós adicionais, até chegar no nó de término de cada ramificação.

Desta forma, inicialmente a árvore de decisão exibiu a seguinte representação:

Classificação de multas aplicadas - Variável Renúncia

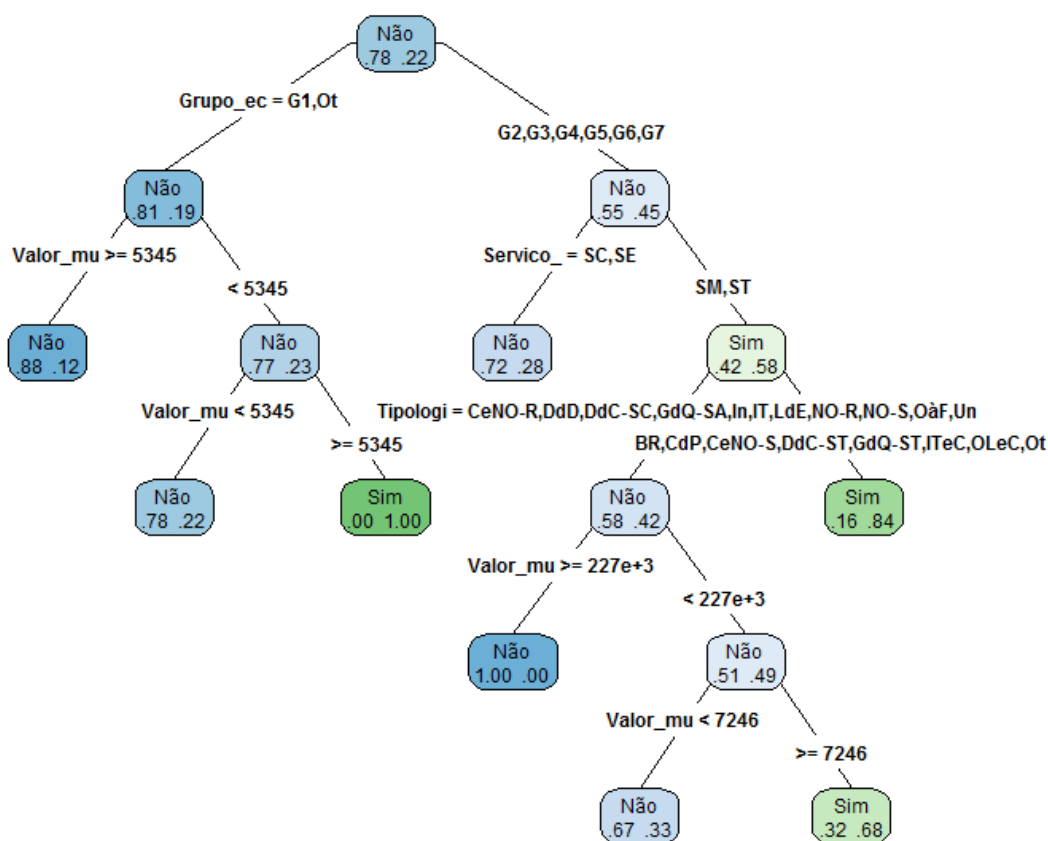


Figura 3 – Árvore de decisão da variável “Renúncia”

Para medir a qualidade deste modelo inicial, foi realizado o teste de validação cruzada, no qual o conjunto de dados foi dividido em 10 partes, sendo que uma parte é utilizada como conjunto de validação enquanto as outras são utilizadas como conjunto de treinamento (*10-fold*). O resultado deste teste está no gráfico abaixo:

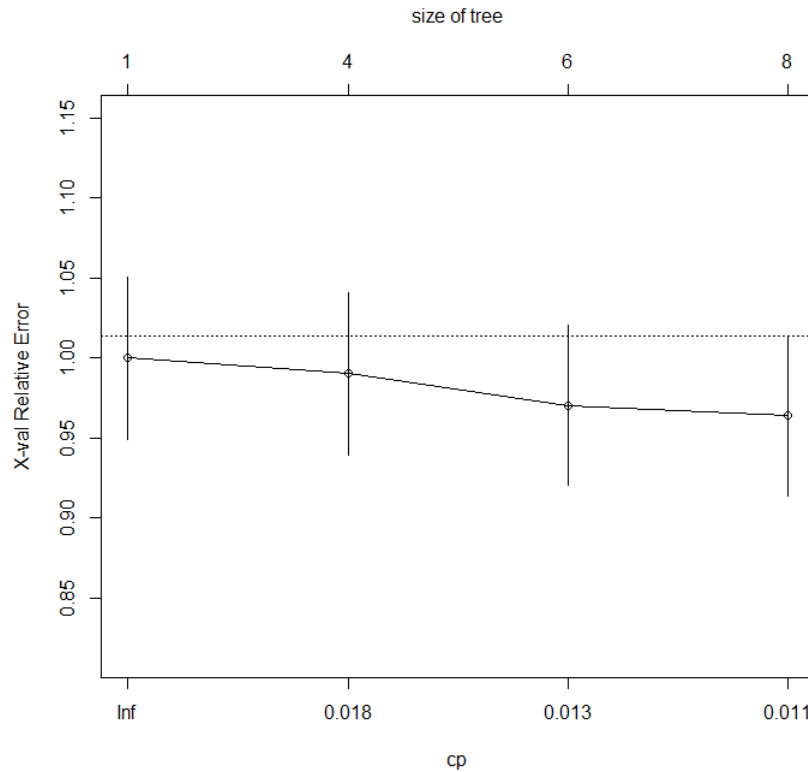


Gráfico 8: Teste de validação cruzada da árvore de decisão da “Renúncia”

Portanto, pela regra *one-standard-deviation* a árvore deve ser podada para 6 nós terminais, resultando em um modelo mais robusto e com menos chance de super-ajuste (*overfitting*). Na figura acima da árvore de decisão, é possível notar que o critério do valor de multa é utilizado diversas vezes. Com a poda da árvore este atributo é utilizado de forma mais generalizada. O resultado deste processo de remoção dos sub-nós de um nó de decisão segue abaixo:

Classificação de multas aplicadas - Variável Renúncia

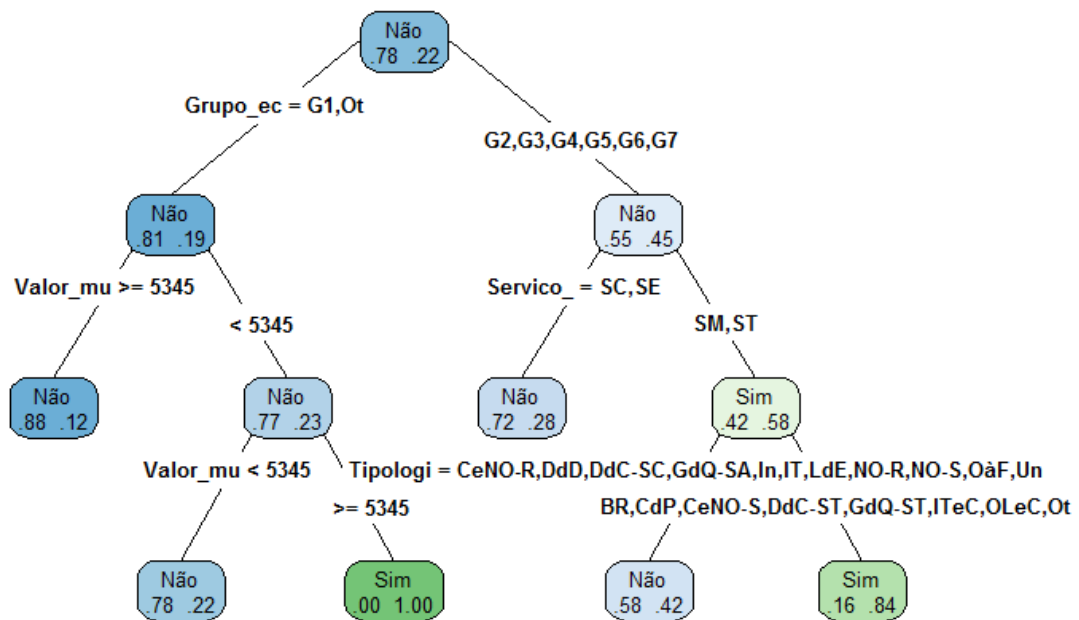


Figura 4 – Árvore de decisão podada da variável “Renúncia”

Comparando a taxa de acerto do modelo nos dois conjuntos (de treinamento e de teste), é possível afirmar que o modelo não apresenta *overfitting*. A taxa de acerto de 80,45% para a base de treino e 79,76% para a base de teste pode ser considerado um bom resultado para um modelo de classificação com alta interpretabilidade.

Comparação de modelos

A título exemplificativo, e de forma a sanar eventuais questionamentos sobre o modelo escolhido, a regressão logística foi testada com a finalidade de comparar a performance dos algoritmos em relação aos atributos escolhidos.

A regressão logística é um modelo supervisionado recomendado para situações em que a variável dependente é de natureza dicotômica ou binária. Dentre algumas de suas vantagens está o alto grau de confiabilidade do algoritmo, além de fornecer os

resultados em termos de probabilidade. É utilizada comumente para avaliações de risco em instituições financeiras e em estudos nas áreas de ciências médicas e sociais, principalmente no campo da econometria.

Desta forma, as mesmas variáveis independentes foram inseridas na construção da regressão logística. Posteriormente, foram aplicados os métodos (utilizando *backward selection* e *stepwise*) para a obtenção de um modelo parcimonioso.

Ao modelo ainda foi possível avaliar a combinação das variáveis “Tipologia de processo”, “Grupo econômico” e “Serviço de telecomunicações” com a variável “Valor de multa”, de forma a testar se estas novas variáveis poderiam ter uma relevância maior diante da variável resposta.

Por fim, foi realizada a estimação dos parâmetros do modelo de regressão logística e avaliação do ajuste obtido.

A taxa de acerto do modelo foi de 80,17% para a base de treino e 79,10% para a base de teste, trazendo resultados muito similares ao modelo de árvore de decisão. Portanto, é necessário realizar uma avaliação mais aprofundada em relação aos dois modelos.

Os resultados dos métodos utilizados, bem como do ajuste aplicado na regressão estão disponíveis no Anexo II deste trabalho.

Avaliação dos modelos

Para avaliação dos modelos estudados, foram consideradas duas métricas de avaliação: a curva ROC (e AUROC) e KS (Kolmogorov-Smirnov).

A curva ROC possui dois parâmetros: taxa de verdadeiro positivo (TPR) e taxa de falso positivo (FPR). É uma curva de probabilidade que mede o grau de separabilidade, pois indica a proporção de acertos do algoritmo. A AUROC apenas resume a curva em um valor único, indicando a área sob a curva.

Em relação aos modelos analisados, a curva ROC (e seu valor em AUROC) da árvore de decisão (com 64,05%) é melhor do que da regressão logística (56,35%), conforme se observa a seguir:

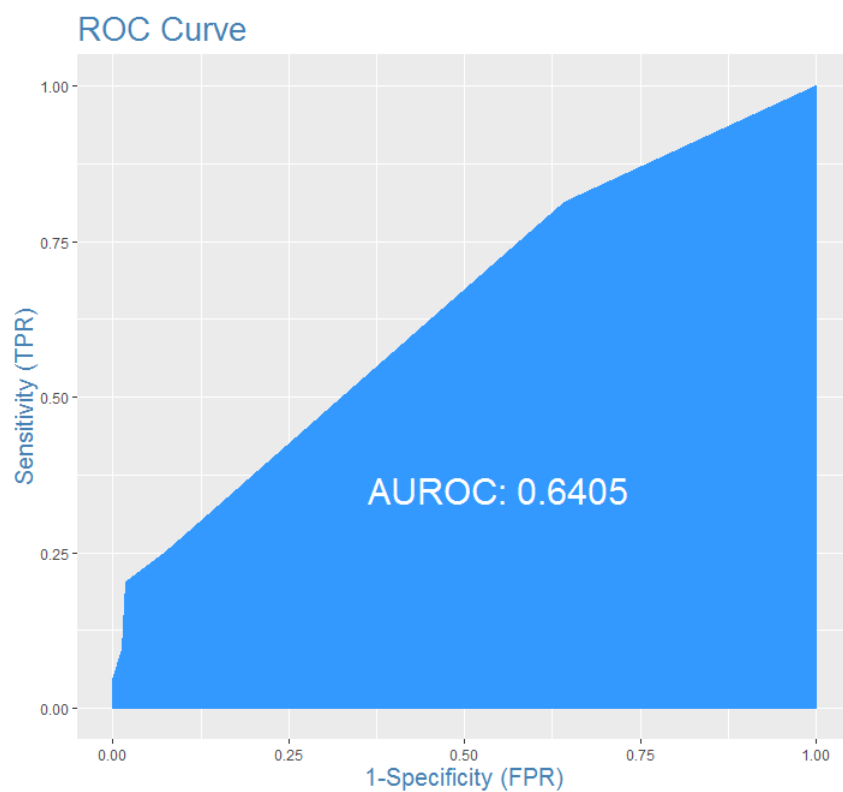


Gráfico 9: Curva ROC da árvore de decisão

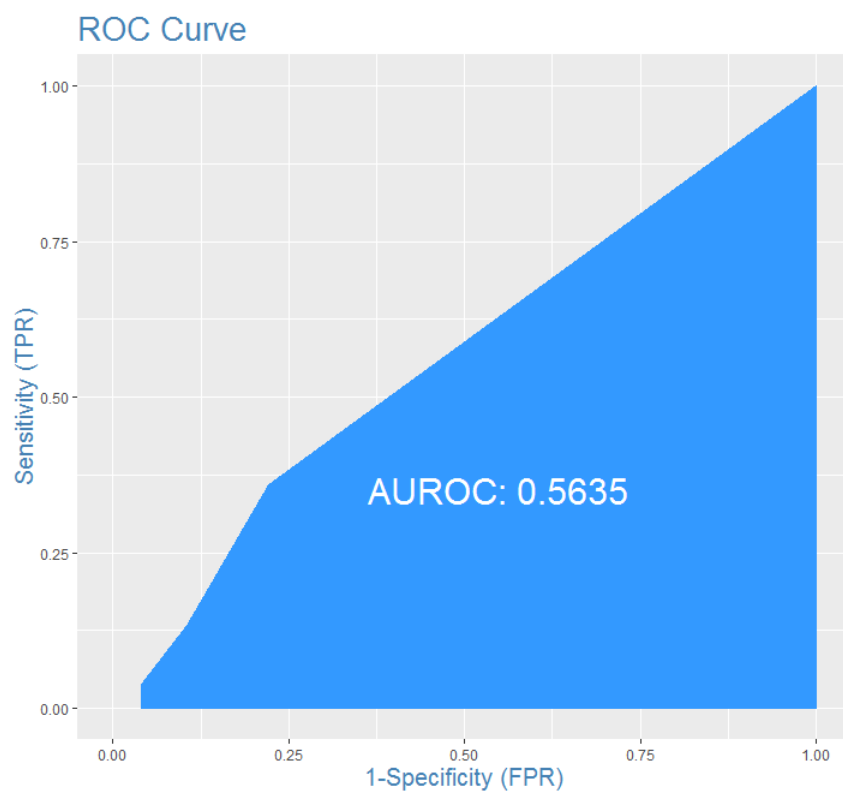


Gráfico 10: Curva ROC da regressão logística

O teste de Kolmogorov-Smirnov (KS) também é uma medida de separabilidade, pois mede a discriminação do modelo estudado.

Abaixo é possível identificar que o modelo árvore de decisão possui percentagens de acerto melhores do que a regressão:

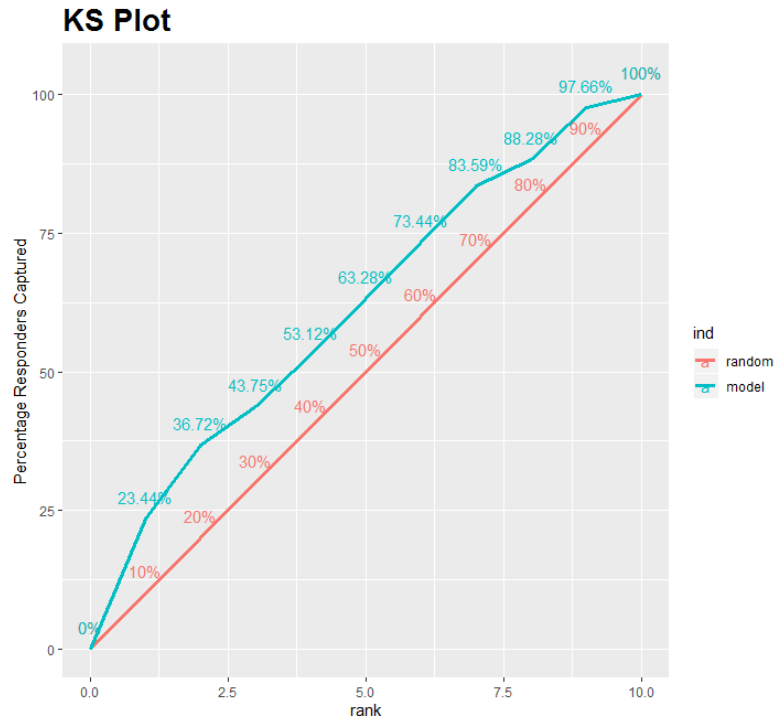


Gráfico 11: Teste KS da árvore de decisão

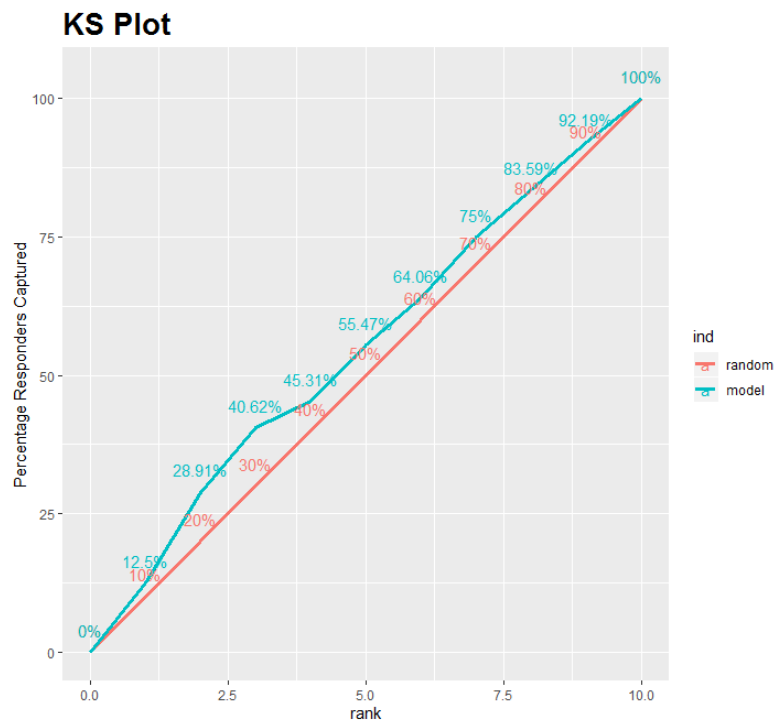


Gráfico 12: Teste KS da regressão logística

Por fim, ambos os modelos foram avaliados de acordo com a taxa de sensibilidade e especificidade, e novamente a árvore de decisão apresentou melhores resultados, conforme se observa a seguir:

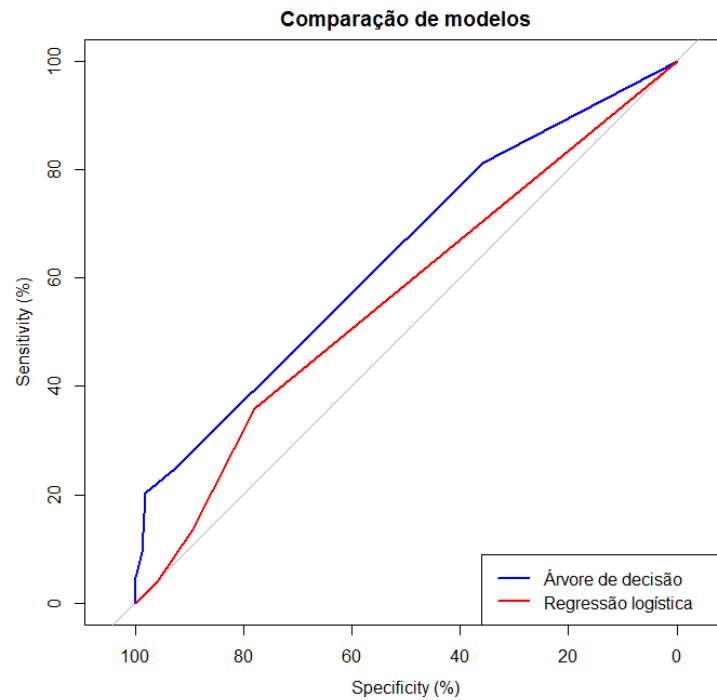


Gráfico 13: Comparação dos modelos pela sensibilidade e especificidade

Como o melhor modelo é a árvore de decisão, calculou-se o intervalo de confiança para o nível de 5% deste modelo de acordo com os parâmetros acima:

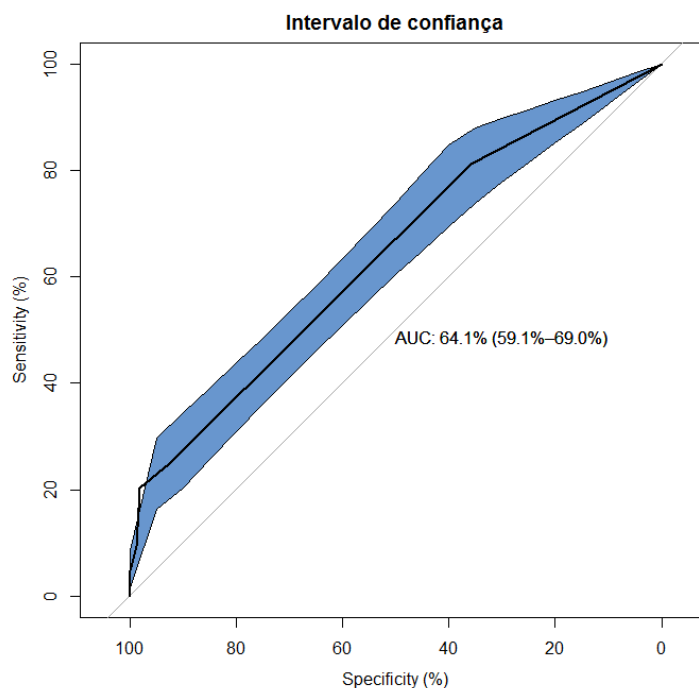


Gráfico 14: Intervalo de confiança do modelo árvore de decisão

Resultados encontrados

O objetivo de qualquer árvore de decisão é criar um modelo viável que preveja o valor de uma variável de destino com base no conjunto de variáveis de entrada (BELL, 2014). Conforme dito anteriormente, uma árvore é uma coleção de elementos chamados nós, dentre os quais um é distinguido como uma raiz, juntamente com uma relação de “paternidade” que impõe uma estrutura hierárquica sobre os nós.

O processo de indução de uma árvore de decisão baseia-se na premissa do “dividir para conquistar”, segregando o espaço definido pelos atributos em subespaços menores, sendo cada subespaço dividido em novos subespaços ou associados a uma classe.

Neste trabalho, o objetivo da utilização do método da árvore de decisão é obter padrões de comportamento dos infratores, particularmente daqueles que optam pela renúncia do seu direito de defesa e consequente pagamento da multa com desconto de 25%, conforme determina a Resolução nº 589, de 7 de maio de 2012, a qual regulamenta a aplicação de Sanções Administrativas da Anatel.

Após realizar a poda da árvore inicial, o algoritmo trouxe como resultado uma árvore com 6 nós terminais. Os resultados que serão analisados abaixo serão baseados neste modelo.

A figura a seguir ilustra o modelo da árvore podada com ênfase nas variáveis. É a mesma árvore de decisão ilustrada na figura 4 acima, porém com informações adicionais.

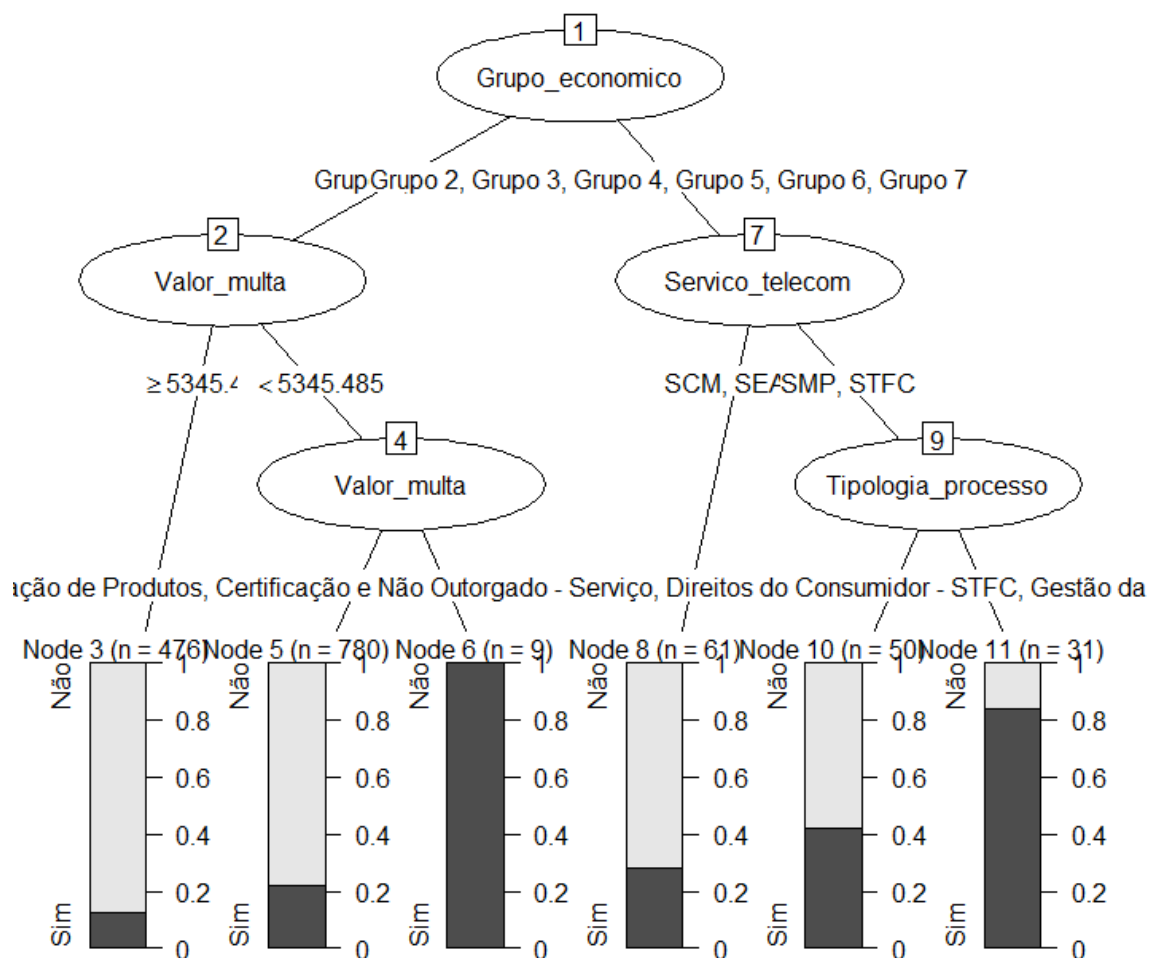


Figura 5 – Árvore de decisão podada da variável “Renúncia”

Os resultados demonstram que a importância das variáveis neste modelo seguiu a seguinte ordem: “Grupo econômico”, “Valor da multa”, “Tipologia de processo”, “Serviço de telecomunicação”. Com a podagem realizada, a variável “Ano de instauração” e “Superintendência” acabaram sendo retiradas do modelo, já que suas importâncias não foram significativas o suficiente. Todas as análises, bem como os resultados de cada divisão e sub-nós podem ser vistos no Anexo II deste trabalho.

Desta forma, é possível inferir pela figura acima que todos os grupos econômicos, com exceção do Grupo 1 e das entidades classificadas em “Outros”, estão inclinados a decidir pela renúncia ou não baseadas no tipo de serviço de telecomunicação infringido. Posteriormente, o critério utilizado seria o tipo de processo de infração.

Por outro lado, o Grupo 1 e as entidades classificadas em “Outros” indicam que o critério de renúncia ou não de seus processos é baseado no valor da multa primeiramente. Em seguida, a tipologia do processo é levada em consideração.

Uma análise mais aprofundada dos dados, bem como a avaliação de outras variáveis podem trazer informações adicionais que contribuam para o modelo deste trabalho. Por exemplo, para a variável “Valor da multa” o modelo indicou um valor de corte na divisão em sub-nós. O valor sugerido pode ser analisado juntamente com outros atributos para avaliar a pertinência de alteração do ato normativo vigente.

Conclusão

Neste trabalho foi explorado o uso de métodos supervisionados baseados em árvores de decisão aplicados na classificação de critérios de tomada de decisão em processos de infração com aplicação de multa pela Anatel.

Além do modelo apresentado, analisou-se um modelo alternativo, baseado na regressão logística, a fim de verificar a aderência do modelo escolhido para os objetivos propostos neste estudo.

Os resultados indicaram uma forte importância de algumas variáveis na opção pela renúncia do direito de defesa dos infratores. Esta avaliação pode subsidiar estudos posteriores na aplicação da fiscalização responsiva na Agência.

Apesar das posteriores críticas e propostas de aprimoramento da teoria, inclusive de um dos próprios autores (BRAITHWAITE, 2007) a fiscalização responsiva pode ser adaptada ao contexto brasileiro. Uma das situações que devem ser ajustadas para este caso é a escalada de ações partindo-se de medidas mais brandas de persuasão conforme o comportamento das empresas. Segundo o autor, grau de reincidência e severidade de danos causados, bem como o nível de risco das infrações devem ser levados em consideração na atuação fiscalizatória.

Um importante passo para isso foi dado com a nova lei das agências reguladoras (Lei nº 13.848/2019), atribuindo a realização de Análise de Impacto Regulatório (AIR), conforme descrito no artigo 6, *in verbis*:

Art. 6º A adoção e as propostas de alteração de atos normativos de interesse geral dos agentes econômicos, consumidores ou usuários dos serviços prestados serão, nos termos de regulamento, precedidas da realização de Análise de Impacto Regulatório (AIR), que conterá informações e dados sobre os possíveis efeitos do ato normativo.

A Análise de Impacto Regulatório – AIR, conforme o Manual de Diretrizes Gerais e Roteiro Analítico Sugerido para Análise de Impacto Regulatório – Diretrizes Gerais AIR da Casa Civil, é o processo sistemático de análise baseado em evidências que busca avaliar, a partir da definição de um problema regulatório, os possíveis impactos das alternativas de ação disponíveis para o alcance dos objetivos pretendidos, tendo como finalidade orientar e subsidiar a tomada de decisão.

Portanto, é notória a ênfase que o poder público vem dando para a análise de dados e para avaliações baseadas em previsão de cenários diferentes para situações que requerem medidas proporcionais ao impacto dos efeitos pretendidos.

Conforme Bencina (2011) abordagens sistemáticas para tomar decisões no setor público estão se tornando muito comum. É um processo de tomada de decisão baseado em sistemas para apoio à tomada de decisão que inclui métodos, modelos e ferramentas, e oferece subsídio de qualidade na tomada de decisão. Uma abordagem como árvores de decisão deve permitir:

- uma integração de numerosas partes interessadas e formação de grupos;
- a percepção de múltiplos problemas e preferências;
- o tratamento de grandes conjuntos de critérios de avaliação;
- uma introdução simples e compreensível ao problema de tomada de decisão e as decisões; e
- uma análise das diferenças nas preferências e na realização de uma pesquisa de concordância das partes interessadas.

O autor vai além, afirmando que a ideia principal de um gráfico orientado a partir de um ponto particular (raiz), o qual se ramifica em uma estrutura conectada de nós e termina em folhas, é usar o gráfico para diferentes propósitos com numerosas abordagens diferentes. Em geral, as árvores de decisão podem ser utilizadas para um dos seguintes fins:

- como uma classificação, ou seja, um cenário, onde as decisões na raiz e nos nós levar a muitos resultados, isto é, as folhas da árvore de decisão;
- como estrutura de avaliação de alternativas, onde a avaliação das folhas com uma agregação dos nós para a raiz resulta em uma avaliação; e
- como estrutura comparativa para encontrar diferenças ou mudanças no estado de a estrutura.

Desta forma, para trabalhos futuros é possível aproveitar os resultados deste estudo a fim de aprofundar no processo de investigação do tema.

O trabalho apresentou alguns indicativos de comportamento dos infratores da Anatel, demonstrando a relevância das variáveis estudadas neste contexto.

Entretanto, tendo em vista se tratar de uma avaliação inédita, o trabalho pode ser melhor aproveitado utilizando-se as informações obtidas de forma a focalizar em objetivos de impacto regulatório específicos.

Conforme dito anteriormente, o modelo indicou que algumas variáveis teriam um peso maior como critério de tomada de decisão para os infratores. Essas variáveis podem ser analisadas em outros modelos e com outros atributos, o que pode resultar em informações valiosas para ações de regulamentação do mercado.

Pode-se também aplicar outras arquiteturas de classificação de padrão, comparando-as com o modelo da árvore de decisão. Esta avaliação pode ser muito importante nas atividades de fiscalização responsiva, na medida que pode auxiliar no cálculo de custo regulatório e principalmente na criação de ações escalonadas de intervenções persuasivas e dissuasivas de regulação.

Por fim, espera-se que esta análise contribua para o aprimoramento da função regulatória da Agência, indicando tendências no comportamento dos infratores e assim contribuindo para a adequação dos atos normativos da Agência. A partir do conhecimento adquirido, o desafio será utiliza-lo de forma adequada para influenciar de maneira eficaz e efetiva em ações de fiscalização e governança.

Lodge e Wegrich (2012) reforçam que a regulação deve ser entendida como um campo de prática contestado, no qual a análise é, em primeiro lugar, compreender a pluralidade de opções e, em segundo, entender os pré-requisitos para uma alternativa capaz de oferecer uma opção mais persuasiva ou aceitável do que outra.

As ferramentas de mineração de dados surgem como um grande aliado no diagnóstico de ações voltadas para a conformidade regulatória e na melhoria da qualidade do serviço público prestado. Assim, espera-se que esses instrumentos de análise de dados possam ser cada vez mais utilizados em benefício do aprimoramento da regulação, contribuindo para a evolução do modelo regulatório brasileiro.

Referências bibliográficas

- BELL, J. Machine Learning: Hands-on for developers and technical professionals. [S.l.]: John Wiley & Sons, 2014.
- BREIMAN, L.; FRIEDMAN, J.; OLSHEN R.; STONE C.; Classification and Regression Trees. London, Chapman & Hall/CRC, 1998.
- JAMES, G.; WITTEN, D.; HASTIE T.; TIBSHIRANI R.; An Introduction to Statistical Learning with Applications in R. James, Witten, Hastie & Tibshirani, 2013.
- KUHN M.; JOHNSON K.; Applied Predictive Modeling, Springer, 2013.
- RAMOS, J.A.P., Árvores de Decisão Aplicadas à Detecção de Fraudes Bancárias, Brasília, 2014, Universidade de Brasília.
- POLLETTINI, J.T.; Avaliação de mecanismos de suporte à tomada de decisão e sua aplicabilidade no auxílio à priorização de casos em regulações de urgências e emergências, Ribeirão Preto, 2016, Faculdade de Medicina de Ribeirão Preto.
- BORGES, F.A.S.; Extração de Características Combinadas com Árvore de Decisão para Detecção e Classificação dos Distúrbios de Qualidade da Energia Elétrica, São Carlos, 2013, Universidade de São Paulo.
- BENCINA, J.; Fuzzy Decisions Trees as a Decision Making Framework in the Public Sector, Slovenia, 2011, Universidade de Ljubljana.
- FERNANDES, C.A.G.; A Reforma do Modelo de Fiscalização do Setor Elétrico Brasileiro, Brasília, 2018, Escola Nacional de Administração Pública.
- CARVALHO M.; Uma Introdução à Análise de Políticas Públicas: Análise Custo-Benefício, Árvores de Decisão e Modelos de Multiatributos, Maringá, 2005, EnANPAD 2005.
- LODGE, Martin; WEGRICH, Kai. Enforcement. In: Managing Regulation – Regulatory Analysis, Politics and Policy. Basingstoke: Palgrave Macmillan, 2012.
- AYRES, I., & BRAITHWAITE, J. Responsive Regulation: Transcending the Deregulation Debate. Oxford: Oxford University Press, 1992.
- BRAITHWAITE, J, MAKKAI, T, BRAITHWAITE, V. Regulating Aged Care: Ritualism and the New Pyramid. Chentelham, UK. Edward Elgar, 2007.

Anexo I – Curriculum resumido da autora

Bacharel em Administração de Empresas pela Universidade de São Paulo – 2005

MBA em Planejamento, Orçamento e Gestão Pública pela Fundação Getúlio Vargas – 2015

Analista-administrativo na Agência Nacional de Telecomunicações – Anatel

Contato: herika@anatel.gov.br

Anexo II – Compilação dos resultados obtidos

```
> # Arvore com a variavel "Renuncia" com dados nao balanceados
> cart <- rpart(Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo + Servico_telecom
+ Grupo_economico + Valor_multa, data = treino)
> prp(cart, main = "Classificação de multas aplicadas - Variável Renúncia", type = 4, extra = 3, box.palette
= "auto", facilen = 2)
> # Teste de validacao cruzada
> printcp(cart)
Classification tree:
rpart(formula = Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo +
Servico_telecom + Grupo_economico + Valor_multa, data = treino)
Variables actually used in tree construction:
[1] Grupo_economico Servico_telecom Tipologia_processo Valor_multa
Root node error: 305/1407 = 0.21677
n= 1407
  CP nsplit rel error  xerror  xstd
1 0.021311  0 1.00000 1.00000 0.050675
2 0.014754  3 0.93115 0.99016 0.050494
3 0.011475  5 0.90164 0.97049 0.050125
4 0.010000  7 0.87869 0.96393 0.050001
> plotcp(cart)
> # Poda da arvore
> pcart<- prune(cart, cp=0.013)
> summary(pcart)
Call:
rpart(formula = Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo +
Servico_telecom + Grupo_economico + Valor_multa, data = treino)
n= 1407
  CP nsplit rel error  xerror  xstd
1 0.02131148  0 1.0000000 1.0000000 0.05067501
2 0.01475410  3 0.9311475 0.9901639 0.05049376
3 0.01300000  5 0.9016393 0.9704918 0.05012519
Variable importance
  Valor_multa  Grupo_economico Tipologia_processo  Servico_telecom  Superintendencia
      35          30          17          14          5
Node number 1: 1407 observations, complexity param=0.02131148
predicted class=Não expected loss=0.2167733 P(node)=1
class counts: 1102 305
probabilities: 0.783 0.217
left son=2 (1265 obs) right son=3 (142 obs)
Primary splits:
  Grupo_economico splits as LRRRRRRL, improve=17.286110, (0 missing)
  Servico_telecom splits as LRRR, improve=10.896490, (0 missing)
```

Tipologia_processo splits as RRLLLLRLRLLLLLLLLLRLRR, improve= 4.949368, (0 missing)
Superintendencia splits as RL, improve= 2.856572, (0 missing)
Valor_multa < 5345.485 to the right, improve= 2.267481, (0 missing)
Surrogate splits:
Tipologia_processo splits as LRLLLLRLLLLLLLLLLLLLLRL, agree=0.905, adj=0.063, (0 split)
Valor_multa < 623420.9 to the left, agree=0.902, adj=0.028, (0 split)
Node number 2: 1265 observations, complexity param=0.0147541
predicted class=Não expected loss=0.1905138 P(node) =0.899076
class counts: 1024 241
probabilities: 0.809 0.191
left son=4 (476 obs) right son=5 (789 obs)
Primary splits:
Valor_multa < 5345.485 to the right, improve=6.76289400, (0 missing)
Servico_telecom splits as LRL, improve=2.97723700, (0 missing)
Tipologia_processo splits as RLRLLLLRLLLLRLRLRL-, improve=1.89965300, (0 missing)
Ano_instauracao splits as RRLLLLR, improve=0.67163270, (0 missing)
Grupo_economico splits as L-----R, improve=0.09055542, (0 missing)
Surrogate splits:
Tipologia_processo splits as RLRLRLLLLRLRRRRRLRL-, agree=0.668, adj=0.118, (0 split)
Servico_telecom splits as RLRL, agree=0.647, adj=0.061, (0 split)
Superintendencia splits as LR, agree=0.631, adj=0.019, (0 split)
Grupo_economico splits as L-----R, agree=0.631, adj=0.019, (0 split)
Ano_instauracao splits as RRRRLRL, agree=0.625, adj=0.002, (0 split)
Node number 3: 142 observations, complexity param=0.02131148
predicted class=Não expected loss=0.4507042 P(node) =0.100924
class counts: 78 64
probabilities: 0.549 0.451
left son=6 (61 obs) right son=7 (81 obs)
Primary splits:
Servico_telecom splits as LLRL, improve=6.3284790, (0 missing)
Valor_multa < 2010 to the left, improve=5.6302830, (0 missing)
Tipologia_processo splits as -RLRLRL-LRLRLRLRLRL, improve=5.5792190, (0 missing)
Ano_instauracao splits as RRRRLRL, improve=2.5221430, (0 missing)
Grupo_economico splits as -LRLRL-, improve=0.4807641, (0 missing)
Surrogate splits:
Tipologia_processo splits as -RRRRRLRLRLRLRLRLRL, agree=0.676, adj=0.246, (0 split)
Valor_multa < 3749.88 to the left, agree=0.655, adj=0.197, (0 split)
Grupo_economico splits as -RRRLRL-, agree=0.585, adj=0.033, (0 split)
Ano_instauracao splits as RRRRLRL, agree=0.577, adj=0.016, (0 split)
Node number 4: 476 observations
predicted class=Não expected loss=0.1239496 P(node) =0.3383085
class counts: 417 59
probabilities: 0.876 0.124
Node number 5: 789 observations, complexity param=0.0147541
predicted class=Não expected loss=0.2306717 P(node) =0.5607676
class counts: 607 182
probabilities: 0.769 0.231
left son=10 (780 obs) right son=11 (9 obs)
Primary splits:
Valor_multa < 5345.435 to the left, improve=10.7765100, (0 missing)
Tipologia_processo splits as R-LLRL--RLLLLLLLLLL-, improve= 3.1686770, (0 missing)
Servico_telecom splits as LRL, improve= 1.3895170, (0 missing)
Ano_instauracao splits as RRLLLLR, improve= 0.5641875, (0 missing)
Superintendencia splits as RL, improve= 0.1042670, (0 missing)
Node number 6: 61 observations
predicted class=Não expected loss=0.2786885 P(node) =0.04335466
class counts: 44 17
probabilities: 0.721 0.279
Node number 7: 81 observations, complexity param=0.02131148
predicted class=Sim expected loss=0.4197531 P(node) =0.0575693

```
class counts: 34 47
probabilities: 0.420 0.580
left son=14 (50 obs) right son=15 (31 obs)
Primary splits:
  Tipologia_processo splits as -RRLRLLR-LRLLRLLL-RLRL, improve=6.7096930, (0 missing)
  Valor_multa < 1950 to the left, improve=3.2990440, (0 missing)
  Ano_instauracao splits as LLRRL-LL-, improve=1.8710760, (0 missing)
  Servico_telecom splits as --LR, improve=0.9679012, (0 missing)
  Grupo_economico splits as -LLRRLR-, improve=0.5094217, (0 missing)
Surrogate splits:
  Superintendencia splits as RL, agree=0.765, adj=0.387, (0 split)
  Servico_telecom splits as --LR, agree=0.691, adj=0.194, (0 split)
  Valor_multa < 586096 to the left, agree=0.691, adj=0.194, (0 split)
Node number 10: 780 observations
predicted class=Não expected loss=0.2217949 P(node) =0.554371
class counts: 607 173
probabilities: 0.778 0.222
Node number 11: 9 observations
predicted class=Sim expected loss=0 P(node) =0.006396588
class counts: 0 9
probabilities: 0.000 1.000
Node number 14: 50 observations
predicted class=Não expected loss=0.42 P(node) =0.0355366
class counts: 29 21
probabilities: 0.580 0.420
Node number 15: 31 observations
predicted class=Sim expected loss=0.1612903 P(node) =0.02203269
class counts: 5 26
probabilities: 0.161 0.839
> # Arvore "podada"
> prp(pcart, main = "Classificação de multas aplicadas - Variável Renúncia",type = 4, extra =
4,box.palette = "auto",faclen = 2)
> plot(as.party(pcart), type="extended")
> # Avaliacao do modelo arvore de decisao
> p.treino <- predict(pcart, type = "class", treino)
> c.treino <- table(treino$Renuncia, p.treino)
> c.treino #matriz de confusao
p.treino
Não Sim
Não 1097 5
Sim 270 35
> sum(diag(prop.table(c.treino)))
[1] 0.8045487
> p.teste <- predict(pcart, type = "class", teste)
> c.teste <- table(teste$Renuncia, p.teste)
> c.teste #matriz de confusao
p.teste
Não Sim
Não 469 6
Sim 116 12
> sum(diag(prop.table(c.teste)))
[1] 0.7976783
>> # Arvore com a variavel "Renuncia" com dados balanceados
> cart_balanceado <- rpart(Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo +
Servico_telecom + Grupo_economico + Valor_multa, data = treino_balanceado)
> prp(cart_balanceado, main = "Classificação de multas aplicadas - Variável Renúncia",type = 4, extra =
4,box.palette = "auto",faclen = 0)
> # Teste de validacao cruzada
> printcp(cart_balanceado)
Classification tree:
```

```
rpart(formula = Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo +
Servico_telecom + Grupo_economico + Valor_multa, data = treino_balanceado)
Variables actually used in tree construction:
[1] Grupo_economico Servico_telecom Tipologia_processo Valor_multa
Root node error: 294/1407 = 0.20896
n= 1407
  CP nsplit rel error  xerror  xstd
1 0.037415  0 1.00000 1.00000 0.051871
2 0.023810  2 0.92517 0.97279 0.051344
3 0.018707  3 0.90136 0.92177 0.050313
4 0.010000  5 0.86395 0.89116 0.049666
> plotcp(cart_balanceado)
> # Poda da arvore
> pcart_balanceado <- prune(cart_balanceado, cp=0.021)
> summary(pcart_balanceado)
Call:
rpart(formula = Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo +
Servico_telecom + Grupo_economico + Valor_multa, data = treino_balanceado)
n= 1407
  CP nsplit rel error  xerror  xstd
1 0.03741497  0 1.0000000 1.0000000 0.05187127
2 0.02380952  2 0.9251701 0.9727891 0.05134421
3 0.02100000  3 0.9013605 0.9217687 0.05031291
Variable importance
  Grupo_economico  Servico_telecom Tipologia_processo  Valor_multa  Ano_instauracao
            41             22             21             14             2
Node number 1: 1407 observations, complexity param=0.03741497
predicted class=Não expected loss=0.2089552 P(node)=1
class counts: 1113 294
probabilities: 0.791 0.209
left son=2 (1288 obs) right son=3 (119 obs)
Primary splits:
  Grupo_economico splits as LRRRLRL, improve=18.958310, (0 missing)
  Tipologia_processo splits as LLRLRLRLRLRLRLRLRLRL, improve= 5.500185, (0 missing)
  Servico_telecom splits as LRRR, improve= 4.844356, (0 missing)
  Ano_instauracao splits as LRRLRL, improve= 4.320161, (0 missing)
  Valor_multa < 5345.485 to the right, improve= 2.662337, (0 missing)
Node number 2: 1288 observations
predicted class=Não expected loss=0.1840062 P(node)=0.9154229
class counts: 1051 237
probabilities: 0.816 0.184
Node number 3: 119 observations, complexity param=0.03741497
predicted class=Não expected loss=0.4789916 P(node)=0.08457711
class counts: 62 57
probabilities: 0.521 0.479
left son=6 (57 obs) right son=7 (62 obs)
Primary splits:
  Servico_telecom splits as LLRR, improve=10.1929200, (0 missing)
  Valor_multa < 4939.915 to the left, improve= 5.9684580, (0 missing)
  Tipologia_processo splits as --RLRL---L-LRRL-RRRRR-, improve= 5.6537820, (0 missing)
  Ano_instauracao splits as RRRRLRL, improve= 2.2235290, (0 missing)
  Grupo_economico splits as -LRL--L-, improve= 0.6054843, (0 missing)
Surrogate splits:
  Valor_multa < 13796.86 to the left, agree=0.739, adj=0.456, (0 split)
  Tipologia_processo splits as --RRRR---L-RLRL-RLRRR-, agree=0.664, adj=0.298, (0 split)
  Ano_instauracao splits as RRRLLLL, agree=0.563, adj=0.088, (0 split)
  Superintendencia splits as LR, agree=0.529, adj=0.018, (0 split)
  Grupo_economico splits as -LRR--R-, agree=0.529, adj=0.018, (0 split)
Node number 6: 57 observations
predicted class=Não expected loss=0.2631579 P(node)=0.04051173
```

```

class counts: 42 15
probabilities: 0.737 0.263
Node number 7: 62 observations, complexity param=0.02380952
predicted class=Sim expected loss=0.3225806 P(node) =0.04406539
class counts: 20 42
probabilities: 0.323 0.677
left son=14 (11 obs) right son=15 (51 obs)
Primary splits:
Tipologia_processo splits as --RLRL-----LLRR-R-RRR-, improve=6.5691450, (0 missing)
Valor_multa < 177191.5 to the right, improve=2.6333160, (0 missing)
Grupo_economico splits as -LRR--R-, improve=1.7723590, (0 missing)
Ano_instauracao splits as RLRLR-, improve=0.9586790, (0 missing)
Superintendencia splits as LR, improve=0.2230368, (0 missing)
Surrogate splits:
Valor_multa < 1617.77 to the left, agree=0.871, adj=0.273, (0 split)
Node number 14: 11 observations
predicted class=Não expected loss=0.1818182 P(node) =0.007818053
class counts: 9 2
probabilities: 0.818 0.182
Node number 15: 51 observations
predicted class=Sim expected loss=0.2156863 P(node) =0.03624733
class counts: 11 40
probabilities: 0.216 0.784
> # Arvore "podada"
> prp(pcart_balanceado, main = "Classificação de multas aplicadas - Variável Renúncia",type = 4, extra =
4,box.palette = "auto",faclen = 0)
> # Avaliacao do modelo arvore de decisao
> p.treino_balanceado <- predict(pcart_balanceado, type = "class", treino_balanceado)
> c.treino_balanceado <- table(treino_balanceado$Renuncia, p.treino_balanceado)
> c.treino_balanceado #matriz de confusao
p.treino_balanceado
Não Sim
Não 1102 11
Sim 254 40
> sum(diag(prop.table(c.treino_balanceado)))
[1] 0.811656
> p.teste_balanceado <- predict(pcart_balanceado, type = "class", teste_balanceado)
> c.teste_balanceado <- table(teste_balanceado$Renuncia, p.teste_balanceado)
> c.teste_balanceado #matriz de confusao
p.teste_balanceado
Não Sim
Não 454 10
Sim 1117 166
> sum(diag(prop.table(c.teste_balanceado)))
[1] 0.3548941
> # Regressao logistica com dados nao balanceados
> rlog <- glm(formula = Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo +
Servico_telecom + Grupo_economico + Valor_multa + Tipologia_processo*Valor_multa +
Servico_telecom*Valor_multa + Grupo_economico*Valor_multa, data = treino, family = binomial())
> summary(rlog2)
Call:
glm(formula = Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo + Servico_telecom
+ Grupo_economico + Valor_multa + Tipologia_processo * Valor_multa + Servico_telecom *
Valor_multa + Grupo_economico * Valor_multa, family = binomial(), data = treino)
Deviance Residuals:
Min 1Q Median 3Q Max
-1.5879 -0.6722 -0.5941 -0.2542 2.6061
Coefficients:
(Intercept) Estimate Std. Error z value Pr(>|z|)
6.356e+01 6.131e+04 0.001 0.999173

```


SuperintendenciaSuperintendencia 2	3.678e-02	3.018e-01	0.122	0.903012
Ano_instauracao2014	-3.652e-02	2.088e-01	-0.175	0.861163
Ano_instauracao2015	-4.400e-02	2.194e-01	-0.201	0.841031
Ano_instauracao2016	-1.925e-01	2.319e-01	-0.830	0.406460
Ano_instauracao2017	-2.695e-01	2.647e-01	-1.018	0.308710
Ano_instauracao2018	-5.420e-01	3.576e-01	-1.515	0.129663
Ano_instauracao2019	5.694e-01	9.749e-01	0.584	0.559197
Tipologia_processoBens Reversíveis	-4.108e+01	8.076e+04	-0.001	0.999594
Tipologia_processoCertificação de Produtos	-6.481e+01	6.131e+04	-0.001	0.999157
Tipologia_processoCertificação e Não Outorgado - Radiofrequência	-6.605e+01	6.131e+04	-0.001	0.999141
Tipologia_processoCertificação e Não Outorgado - Serviço	-6.244e+01	6.131e+04	-0.001	0.999187
Tipologia_processoDescumprimento de Determinação	-6.498e+01	6.131e+04	-0.001	0.999154
Tipologia_processoDireitos do Consumidor - SCM	-6.510e+01	6.131e+04	-0.001	0.999153
Tipologia_processoDireitos do Consumidor - STFC	-6.502e+01	6.131e+04	-0.001	0.999154
Tipologia_processoDisponibilidade e Funcionamento de TUP	-8.844e+01	8.245e+04	-0.001	0.999144
Tipologia_processoGestão da Qualidade - SeAC	-6.516e+01	6.131e+04	-0.001	0.999152
Tipologia_processoGestão da Qualidade - STFC	-8.535e+01	6.190e+04	-0.001	0.998900
Tipologia_processoInterconexão	-6.550e+01	6.131e+04	-0.001	0.999148
Tipologia_processoIrregularidade Técnica	-6.553e+01	6.131e+04	-0.001	0.999147
Tipologia_processoIrregularidade Técnica e Certificação	-6.384e+01	6.131e+04	-0.001	0.999169
Tipologia_processoLicenciamento de Estação	-6.519e+01	6.131e+04	-0.001	0.999152
Tipologia_processoNão Outorgado - Radiofrequência	-6.499e+01	6.131e+04	-0.001	0.999154
Tipologia_processoNão Outorgado - Serviço	-6.499e+01	6.131e+04	-0.001	0.999154
Tipologia_processoNão Outorgado - Serviço e Radiofrequência	-6.576e+01	6.131e+04	-0.001	0.999144
Tipologia_processoObrigações Legais e Contratuais	-6.427e+01	6.131e+04	-0.001	0.999164
Tipologia_processoObstrução à Fiscalização	-6.478e+01	6.131e+04	-0.001	0.999157
Tipologia_processoOutros	-6.487e+01	6.131e+04	-0.001	0.999156
Tipologia_processoUniversalização	3.654e+01	6.137e+04	0.001	0.999525
Servico_telecomSEAC	1.094e+00	3.870e-01	2.827	0.004705 **
Servico_telecomSMP	6.527e-01	2.103e-01	3.104	0.001908 **
Servico_telecomSTFC	1.316e+00	3.591e-01	3.666	0.000247 ***
Grupo_economicoGrupo 2	5.459e-01	5.655e-01	0.965	0.334312
Grupo_economicoGrupo 3	1.166e+00	6.063e-01	1.922	0.054566 .
Grupo_economicoGrupo 4	6.868e-01	6.220e-01	1.104	0.269539
Grupo_economicoGrupo 5	-1.343e-01	1.357e+00	-0.099	0.921138
Grupo_economicoGrupo 6	2.943e-01	1.035e+00	0.284	0.776079
Grupo_economicoGrupo 7	-2.116e+02	5.285e+05	0.000	0.999681
Grupo_economicoOutros	-4.937e-02	4.651e-01	-0.106	0.915470
Valor_multa	-3.528e-02	3.110e+01	-0.001	0.999095
Tipologia_processoBens Reversíveis:Valor_multa	3.525e-02	3.110e+01	0.001	0.999096
Tipologia_processoCertificação de Produtos:Valor_multa	3.526e-02	3.110e+01	0.001	0.999095
Tipologia_processoCertificação e Não Outorgado - Radiofrequência:Valor_multa	3.535e-02	3.110e+01	0.001	0.999093
Tipologia_processoCertificação e Não Outorgado - Serviço:Valor_multa	3.464e-02	3.110e+01	0.001	0.999111
Tipologia_processoDescumprimento de Determinação:Valor_multa	3.518e-02	3.110e+01	0.001	0.999098
Tipologia_processoDireitos do Consumidor - SCM:Valor_multa	3.524e-02	3.110e+01	0.001	0.999096
Tipologia_processoDireitos do Consumidor - STFC:Valor_multa	3.523e-02	3.110e+01	0.001	0.999096
Tipologia_processoDisponibilidade e Funcionamento de TUP:Valor_multa	3.530e-02	3.113e+01	0.001	0.999095
Tipologia_processoGestão da Qualidade - SeAC:Valor_multa	3.526e-02	3.110e+01	0.001	0.999095
Tipologia_processoGestão da Qualidade - STFC:Valor_multa	3.533e-02	3.110e+01	0.001	0.999094
Tipologia_processoInterconexão:Valor_multa	3.524e-02	3.110e+01	0.001	0.999096
Tipologia_processoIrregularidade Técnica:Valor_multa	3.527e-02	3.110e+01	0.001	0.999095
Tipologia_processoIrregularidade Técnica e Certificação:Valor_multa	3.484e-02	3.110e+01	0.001	0.999106

Tipologia_processo	Valor_multa	3.530e-02	3.110e+01	0.001	0.999094
Licenciamento de Estação:Valor_multa					
Não Outorgado - Radiofrequência:Valor_multa		3.527e-02	3.110e+01	0.001	0.999095
Não Outorgado - Serviço:Valor_multa	3.525e-02	3.110e+01	0.001	0.999096	
Não Outorgado - Serviço e Radiofrequência:Valor_multa	3.543e-02	3.110e+01	0.001	0.999091	
Obrigações Legais e Contratuais:Valor_multa	3.522e-02	3.110e+01	0.001	0.999097	
Obstrução à Fiscalização:Valor_multa	3.525e-02	3.110e+01	0.001	0.999096	
Outros:Valor_multa	3.525e-02	3.110e+01	0.001	0.999096	
Universalização:Valor_multa	5.946e-03	3.111e+01	0.000	0.999847	
SEAC:Valor_multa	-2.979e-05	2.126e-05	-1.401	0.161111	
SMP:Valor_multa	-2.548e-05	1.121e-05	-2.272	0.023098 *	
STFC:Valor_multa	-2.885e-05	1.123e-05	-2.569	0.010211 *	
Grupo 2:Valor_multa	5.259e-05	2.743e-05	1.917	0.055218 .	
Grupo 3:Valor_multa	5.023e-05	2.753e-05	1.824	0.068136 .	
Grupo 4:Valor_multa	5.583e-05	2.749e-05	2.031	0.042286 *	
Grupo 5:Valor_multa	4.889e-05	6.769e-05	0.722	0.470167	
Grupo 6:Valor_multa	3.924e-05	4.366e-05	0.899	0.368821	
Grupo 7:Valor_multa	1.731e-03	4.266e+00	0.000	0.999676	
Outros:Valor_multa	2.430e-05	2.784e-05	0.873	0.382818	

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 1471.1 on 1406 degrees of freedom
Residual deviance: 1320.4 on 1336 degrees of freedom
AIC: 1462.4
Number of Fisher Scoring iterations: 22
> step(rlog, trace = 1, direction = c("backward"))
Start: AIC=1462.35
Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo + Servico_telecom +
Grupo_economico + Valor_multa + Tipologia_processo * Valor_multa + Servico_telecom * Valor_multa
+ Grupo_economico * Valor_multa
Df Deviance AIC
- Ano_instauracao          6  1324.2 1454.2
- Superintendencia         1  1320.4 1460.4
<none>                     1320.3 1462.3
- Tipologia_processo:Valor_multa 21  1366.1 1466.1
- Servico_telecom:Valor_multa   3  1332.1 1468.1
- Grupo_economico:Valor_multa  7  1345.7 1473.7
Step: AIC=1454.17
Renuncia ~ Superintendencia + Tipologia_processo + Servico_telecom + Grupo_economico +
Valor_multa + Tipologia_processo:Valor_multa + Servico_telecom:Valor_multa +
Grupo_economico:Valor_multa
Df Deviance AIC
- Superintendencia          1  1324.2 1452.2
<none>                     1324.2 1454.2
- Tipologia_processo:Valor_multa 21  1369.8 1457.8
- Servico_telecom:Valor_multa   3  1335.8 1459.8
- Grupo_economico:Valor_multa  7  1349.7 1465.7
Step: AIC=1452.17
Renuncia ~ Tipologia_processo + Servico_telecom + Grupo_economico + Valor_multa +
Tipologia_processo:Valor_multa + Servico_telecom:Valor_multa + Grupo_economico:Valor_multa
Df Deviance AIC
<none>                     1324.2 1452.2
- Tipologia_processo:Valor_multa 21  1369.8 1455.8
- Servico_telecom:Valor_multa   3  1335.8 1457.8
- Grupo_economico:Valor_multa  7  1349.7 1463.7
Call: glm(formula = Renuncia ~ Tipologia_processo + Servico_telecom + Grupo_economico +
Valor_multa + Tipologia_processo:Valor_multa + Servico_telecom:Valor_multa +
Grupo_economico:Valor_multa, family = binomial(), data = treino)

```

Coefficients:

	(Intercept)	6.324e+01
	Tipologia_processoBens Reversíveis	-4.084e+01
	Tipologia_processoCertificação de Produtos	-6.459e+01
Tipologia_processo	Certificação e Não Outorgado - Radiofrequência	-6.565e+01
	Certificação e Não Outorgado - Serviço	-6.213e+01
	Descumprimento de Determinação	-6.461e+01
	Direitos do Consumidor - SCM	-6.476e+01
	Direitos do Consumidor - STFC	-6.476e+01
Tipologia_processo	Disponibilidade e Funcionamento de TUP	-8.811e+01
	Gestão da Qualidade - SeAC	-6.484e+01
	Gestão da Qualidade - STFC	-8.494e+01
	processoInterconexão	-6.522e+01
	processoIrregularidade Técnica	-6.520e+01
Tipologia_processo	Irregularidade Técnica e Certificação	-6.362e+01
	processoLicenciamento de Estação	-6.489e+01
Tipologia_processo	Não Outorgado - Radiofrequência	-6.476e+01
	Não Outorgado - Serviço	-6.470e+01
Tipologia_processo	Não Outorgado - Serviço e Radiofrequência	-6.541e+01
	processoObrigações Legais e Contratuais	-6.393e+01
	processoObstrução à Fiscalização	-6.466e+01
	processoOutros	-6.450e+01
	processoUniversalização	3.690e+01
	Servico_telecomSEAC	1.072e+00
	Servico_telecomSMP	6.659e-01
	Servico_telecomSTFC	1.298e+00
	Grupo_economicoGrupo 2	5.330e-01
	Grupo_economicoGrupo 3	1.088e+00
	Grupo_economicoGrupo 4	6.163e-01
	Grupo_economicoGrupo 5	-1.835e-01
	Grupo_economicoGrupo 6	2.564e-01
	Grupo_economicoGrupo 7	-2.127e+02
	Grupo_economicoOutros	-1.029e-01
	Valor_multa	-3.511e-02
	Tipologia_processoBens Reversíveis:Valor_multa	3.508e-02
	Tipologia_processoCertificação de Produtos:Valor_multa	3.510e-02
Tipologia_processo	Certificação e Não Outorgado - Radiofrequência:Valor_multa	3.516e-02
	Certificação e Não Outorgado - Serviço:Valor_multa	3.446e-02
	Descumprimento de Determinação:Valor_multa	3.501e-02
	Direitos do Consumidor - SCM:Valor_multa	3.507e-02
	Direitos do Consumidor - STFC:Valor_multa	3.509e-02
Tipologia_processo	Disponibilidade e Funcionamento de TUP:Valor_multa	3.514e-02
	Gestão da Qualidade - SeAC:Valor_multa	3.510e-02
	Gestão da Qualidade - STFC:Valor_multa	3.516e-02
	processoInterconexão:Valor_multa	3.508e-02
	processoIrregularidade Técnica:Valor_multa	3.511e-02
Tipologia_processo	Irregularidade Técnica e Certificação:Valor_multa	3.471e-02
	processoLicenciamento de Estação:Valor_multa	3.514e-02
Tipologia_processo	Não Outorgado - Radiofrequência:Valor_multa	3.510e-02
	Não Outorgado - Serviço:Valor_multa	3.509e-02
Tipologia_processo	Não Outorgado - Serviço e Radiofrequência:Valor_multa	3.526e-02
	processoObrigações Legais e Contratuais:Valor_multa	3.505e-02
	processoObstrução à Fiscalização:Valor_multa	3.508e-02
	processoOutros:Valor_multa	3.509e-02
	processoUniversalização:Valor_multa	5.774e-03
	Servico_telecomSEAC:Valor_multa	-3.000e-05
	Servico_telecomSMP:Valor_multa	-2.524e-05
	Servico_telecomSTFC:Valor_multa	-2.856e-05
	Grupo_economicoGrupo 2:Valor_multa	5.109e-05
	Grupo_economicoGrupo 3:Valor_multa	4.879e-05

```
Grupo_economicoGrupo 4:Valor_multa 5.441e-05
Grupo_economicoGrupo 5:Valor_multa 4.812e-05
Grupo_economicoGrupo 6:Valor_multa 3.821e-05
Grupo_economicoGrupo 7:Valor_multa 1.739e-03
Grupo_economicoOutros:Valor_multa 2.306e-05
Degrees of Freedom: 1406 Total (i.e. Null); 1343 Residual
Null Deviance: 1471
Residual Deviance: 1324 AIC: 1452
There were 14 warnings (use warnings() to see them)
> step(rlog, trace = 1, direction = c("both"))
Start: AIC=1462.35
Renuncia ~ Superintendencia + Ano_instauracao + Tipologia_processo + Servico_telecom +
Grupo_economico + Valor_multa + Tipologia_processo * Valor_multa + Servico_telecom * Valor_multa
+ Grupo_economico * Valor_multa
Df Deviance AIC
- Ano_instauracao 6 1324.2 1454.2
- Superintendencia 1 1320.4 1460.4
<none> 1320.3 1462.3
- Tipologia_processo:Valor_multa 21 1366.1 1466.1
- Servico_telecom:Valor_multa 3 1332.1 1468.1
- Grupo_economico:Valor_multa 7 1345.7 1473.7
Step: AIC=1454.17
Renuncia ~ Superintendencia + Tipologia_processo + Servico_telecom + Grupo_economico +
Valor_multa + Tipologia_processo:Valor_multa + Servico_telecom:Valor_multa +
Grupo_economico:Valor_multa
Df Deviance AIC
- Superintendencia 1 1324.2 1452.2
<none> 1324.2 1454.2
- Tipologia_processo:Valor_multa 21 1369.8 1457.8
- Servico_telecom:Valor_multa 3 1335.8 1459.8
+ Ano_instauracao 6 1320.3 1462.3
- Grupo_economico:Valor_multa 7 1349.7 1465.7
Step: AIC=1452.17
Renuncia ~ Tipologia_processo + Servico_telecom + Grupo_economico + Valor_multa +
Tipologia_processo:Valor_multa + Servico_telecom:Valor_multa + Grupo_economico:Valor_multa
Df Deviance AIC
<none> 1324.2 1452.2
+ Superintendencia 1 1324.2 1454.2
- Tipologia_processo:Valor_multa 21 1369.8 1455.8
- Servico_telecom:Valor_multa 3 1335.8 1457.8
+ Ano_instauracao 6 1320.4 1460.4
- Grupo_economico:Valor_multa 7 1349.7 1463.7
Call: glm(formula = Renuncia ~ Tipologia_processo + Servico_telecom + Grupo_economico +
Valor_multa + Tipologia_processo:Valor_multa + Servico_telecom:Valor_multa +
Grupo_economico:Valor_multa, family = binomial(), data = treino)
Coefficients:
(Intercept) 6.324e+01
Tipologia_processoBens Reversíveis -4.084e+01
Tipologia_processoCertificação de Produtos -6.459e+01
Tipologia_processoCertificação e Não Outorgado - Radiofrequência -6.565e+01
Tipologia_processoCertificação e Não Outorgado - Serviço -6.213e+01
Tipologia_processoDescumprimento de Determinação -6.461e+01
Tipologia_processoDireitos do Consumidor - SCM -6.476e+01
Tipologia_processoDireitos do Consumidor - STFC -6.476e+01
Tipologia_processoDisponibilidade e Funcionamento de TUP -8.811e+01
Tipologia_processoGestão da Qualidade - SeAC -6.484e+01
Tipologia_processoGestão da Qualidade - STFC -8.494e+01
Tipologia_processoInterconexão -6.522e+01
Tipologia_processoIrregularidade Técnica -6.520e+01
Tipologia_processoIrregularidade Técnica e Certificação -6.362e+01
```

```

Tipologia_processoLicenciamento de Estação -6.489e+01
Tipologia_processoNão Outorgado - Radiofrequência -6.476e+01
Tipologia_processoNão Outorgado - Serviço -6.470e+01
Tipologia_processoNão Outorgado - Serviço e Radiofrequência -6.541e+01
Tipologia_processoObrigações Legais e Contratuais -6.393e+01
Tipologia_processoObstrução à Fiscalização -6.466e+01
Tipologia_processoOutros -6.450e+01
Tipologia_processoUniversalização 3.690e+01
Servico_telecomSEAC 1.072e+00
Servico_telecomSMP 6.659e-01
Servico_telecomSTFC 1.298e+00
Grupo_economicoGrupo 2 5.330e-01
Grupo_economicoGrupo 3 1.088e+00
Grupo_economicoGrupo 4 6.163e-01
Grupo_economicoGrupo 5 -1.835e-01
Grupo_economicoGrupo 6 2.564e-01
Grupo_economicoGrupo 7 -2.127e+02
Grupo_economicoOutros -1.029e-01
Valor_multa -3.511e-02
Tipologia_processoBens Reversíveis:Valor_multa 3.508e-02
Tipologia_processoCertificação de Produtos:Valor_multa 3.510e-02
Tipologia_processoCertificação e Não Outorgado - Radiofrequência:Valor_multa 3.516e-02
Tipologia_processoCertificação e Não Outorgado - Serviço:Valor_multa 3.446e-02
Tipologia_processoDescumprimento de Determinação:Valor_multa 3.501e-02
Tipologia_processoDireitos do Consumidor - SCM:Valor_multa 3.507e-02
Tipologia_processoDireitos do Consumidor - STFC:Valor_multa 3.509e-02
Tipologia_processoDisponibilidade e Funcionamento de TUP:Valor_multa 3.514e-02
Tipologia_processoGestão da Qualidade - SeAC:Valor_multa 3.510e-02
Tipologia_processoGestão da Qualidade - STFC:Valor_multa 3.516e-02
Tipologia_processoInterconexão:Valor_multa 3.508e-02
Tipologia_processoIrregularidade Técnica:Valor_multa 3.511e-02
Tipologia_processoIrregularidade Técnica e Certificação:Valor_multa 3.471e-02
Tipologia_processoLicenciamento de Estação:Valor_multa 3.514e-02
Tipologia_processoNão Outorgado - Radiofrequência:Valor_multa 3.510e-02
Tipologia_processoNão Outorgado - Serviço:Valor_multa 3.509e-02
Tipologia_processoNão Outorgado - Serviço e Radiofrequência:Valor_multa 3.526e-02
Tipologia_processoObrigações Legais e Contratuais:Valor_multa 3.505e-02
Tipologia_processoObstrução à Fiscalização:Valor_multa 3.508e-02
Tipologia_processoOutros:Valor_multa 3.509e-02
Tipologia_processoUniversalização:Valor_multa 5.774e-03
Servico_telecomSEAC:Valor_multa -3.000e-05
Servico_telecomSMP:Valor_multa -2.524e-05
Servico_telecomSTFC:Valor_multa -2.856e-05
Grupo_economicoGrupo 2:Valor_multa 5.109e-05
Grupo_economicoGrupo 3:Valor_multa 4.879e-05
Grupo_economicoGrupo 4:Valor_multa 5.441e-05
Grupo_economicoGrupo 5:Valor_multa 4.812e-05
Grupo_economicoGrupo 6:Valor_multa 3.821e-05
Grupo_economicoGrupo 7:Valor_multa 1.739e-03
Grupo_economicoOutros:Valor_multa 2.306e-05
Degrees of Freedom: 1406 Total (i.e. Null); 1343 Residual
Null Deviance: 1471
Residual Deviance: 1324 AIC: 1452
> rlog2 <- glm(formula = Renuncia ~ Tipologia_processo + Servico_telecom + Grupo_economico +
Valor_multa + Tipologia_processo*Valor_multa + Servico_telecom*Valor_multa +
Grupo_economico*Valor_multa, family = binomial(), data = treino)
> summary(rlog2)
Call:

```

```
glm(formula = Renuncia ~ Tipologia_processo + Servico_telecom + Grupo_economico + Valor_multa +
Tipologia_processo * Valor_multa + Servico_telecom * Valor_multa + Grupo_economico *
Valor_multa, family = binomial(), data = treino)
```

Deviance Residuals:

```
Min    1Q  Median    3Q   Max
-1.5581 -0.6652 -0.6159 -0.2524  2.5070
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.324e+01	6.182e+04	0.001	0.999184
Tipologia_processoBens Reversíveis	-4.084e+01	8.088e+04	-0.001	0.999597
Tipologia_processoCertificação de Produtos	-6.459e+01	6.182e+04	-0.001	0.999166
Tipologia_processoCertificação e Não Outorgado - Radiofrequência	-6.565e+01	6.182e+04	-0.001	0.999153
Tipologia_processoCertificação e Não Outorgado - Serviço	-6.213e+01	6.182e+04	-0.001	0.999198
Tipologia_processoDescumprimento de Determinação	-6.461e+01	6.182e+04	-0.001	0.999166
Tipologia_processoDireitos do Consumidor - SCM	-6.476e+01	6.182e+04	-0.001	0.999164
Tipologia_processoDireitos do Consumidor - STFC	-6.476e+01	6.182e+04	-0.001	0.999164
Tipologia_processoDisponibilidade e Funcionamento de TUP	-8.811e+01	8.283e+04	-0.001	0.999151
Tipologia_processoGestão da Qualidade - SeAC	-6.484e+01	6.182e+04	-0.001	0.999163
Tipologia_processoGestão da Qualidade - STFC	-8.494e+01	6.237e+04	-0.001	0.998913
Tipologia_processoInterconexão	-6.522e+01	6.182e+04	-0.001	0.999158
Tipologia_processoIrregularidade Técnica	-6.520e+01	6.182e+04	-0.001	0.999159
Tipologia_processoIrregularidade Técnica e Certificação	-6.362e+01	6.182e+04	-0.001	0.999179
Tipologia_processoLicenciamento de Estação	-6.489e+01	6.182e+04	-0.001	0.999163
Tipologia_processoNão Outorgado - Radiofrequência	-6.476e+01	6.182e+04	-0.001	0.999164
Tipologia_processoNão Outorgado - Serviço	-6.470e+01	6.182e+04	-0.001	0.999165
Tipologia_processoNão Outorgado - Serviço e Radiofrequência	-6.541e+01	6.182e+04	-0.001	0.999156
Tipologia_processoObrigações Legais e Contratuais	-6.393e+01	6.182e+04	-0.001	0.999175
Tipologia_processoObstrução à Fiscalização	-6.466e+01	6.182e+04	-0.001	0.999165
Tipologia_processoOutros	-6.450e+01	6.182e+04	-0.001	0.999168
Tipologia_processoUniversalização	3.690e+01	6.188e+04	0.001	0.999524
Servico_telecomSEAC	1.072e+00	3.773e-01	2.841	0.004498 **
Servico_telecomSMP	6.659e-01	2.091e-01	3.184	0.001451 **
Servico_telecomSTFC	1.298e+00	3.572e-01	3.635	0.000278 ***
Grupo_economicoGrupo 2	5.330e-01	5.615e-01	0.949	0.342501
Grupo_economicoGrupo 3	1.088e+00	6.011e-01	1.810	0.070236 .
Grupo_economicoGrupo 4	6.163e-01	6.143e-01	1.003	0.315741
Grupo_economicoGrupo 5	-1.835e-01	1.351e+00	-0.136	0.891965
Grupo_economicoGrupo 6	2.564e-01	1.014e+00	0.253	0.800388
Grupo_economicoGrupo 7	-2.127e+02	5.285e+05	0.000	0.999679
Grupo_economicoOutros	-1.029e-01	4.570e-01	-0.225	0.821892
Valor_multa	-3.511e-02	3.133e+01	-0.001	0.999106
Tipologia_processoBens Reversíveis:Valor_multa	3.508e-02	3.133e+01	0.001	0.999106
Tipologia_processoCertificação de Produtos:Valor_multa	3.510e-02	3.133e+01	0.001	0.999106
Tipologia_processoCertificação e Não Outorgado - Radiofrequência:Valor_multa	3.516e-02	3.133e+01	0.001	0.999104
Tipologia_processoCertificação e Não Outorgado - Serviço:Valor_multa	3.446e-02	3.133e+01	0.001	0.999122
Tipologia_processoDescumprimento de Determinação:Valor_multa	3.501e-02	3.133e+01	0.001	0.999108
Tipologia_processoDireitos do Consumidor - SCM:Valor_multa	3.507e-02	3.133e+01	0.001	0.999107
Tipologia_processoDireitos do Consumidor - STFC:Valor_multa	3.509e-02	3.133e+01	0.001	0.999106
Tipologia_processoDisponibilidade e Funcionamento de TUP:Valor_multa	3.514e-02	3.136e+01	0.001	0.999106
Tipologia_processoGestão da Qualidade - SeAC:Valor_multa	3.510e-02	3.133e+01	0.001	0.999106
Tipologia_processoGestão da Qualidade - STFC:Valor_multa	3.516e-02	3.133e+01	0.001	0.999104
Tipologia_processoInterconexão:Valor_multa	3.508e-02	3.133e+01	0.001	0.999107
Tipologia_processoIrregularidade Técnica:Valor_multa	3.511e-02	3.133e+01	0.001	0.999106

Tipologia_processo	Irregularidade Técnica e Certificação:Valor_multa	3.471e-02	3.133e+01	0.001	0.999116
Tipologia_processo	Licenciamento de Estação:Valor_multa	3.514e-02	3.133e+01	0.001	0.999105
Tipologia_processo	Não Outorgado - Radiofrequência:Valor_multa	3.510e-02	3.133e+01	0.001	0.999106
Tipologia_processo	Não Outorgado - Serviço:Valor_multa	3.509e-02	3.133e+01	0.001	0.999106
Tipologia_processo	Não Outorgado - Serviço e Radiofrequência:Valor_multa	3.526e-02	3.133e+01	0.001	0.999102
Tipologia_processo	Obrigações Legais e Contratuais:Valor_multa	3.505e-02	3.133e+01	0.001	0.999107
Tipologia_processo	Obstrução à Fiscalização:Valor_multa	3.508e-02	3.133e+01	0.001	0.999106
Tipologia_processo	Outros:Valor_multa	3.509e-02	3.133e+01	0.001	0.999106
Tipologia_processo	Universalização:Valor_multa	5.774e-03	3.134e+01	0.000	0.999853
Servico_telecom	SEAC:Valor_multa	-3.000e-05	2.098e-05	-1.430	0.152758
Servico_telecom	SMP:Valor_multa	-2.524e-05	1.116e-05	-2.262	0.023720 *
Servico_telecom	STFC:Valor_multa	-2.856e-05	1.119e-05	-2.552	0.010707 *
Grupo_economico	Grupo 2:Valor_multa	5.109e-05	2.701e-05	1.891	0.058573 .
Grupo_economico	Grupo 3:Valor_multa	4.879e-05	2.712e-05	1.799	0.071993 .
Grupo_economico	Grupo 4:Valor_multa	5.441e-05	2.708e-05	2.009	0.044530 *
Grupo_economico	Grupo 5:Valor_multa	4.812e-05	6.709e-05	0.717	0.473235
Grupo_economico	Grupo 6:Valor_multa	3.821e-05	4.048e-05	0.944	0.345211
Grupo_economico	Grupo 7:Valor_multa	1.739e-03	4.266e+00	0.000	0.999675
Grupo_economico	Outros:Valor_multa	2.306e-05	2.751e-05	0.838	0.401824

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 1471.1 on 1406 degrees of freedom
Residual deviance: 1324.2 on 1343 degrees of freedom
AIC: 1452.2
Number of Fisher Scoring iterations: 22
> # Avaliacao do modelo de regressao logistica
> probabilidades.treino <- predict(rlog2, type = "response", treino)
> predicted.classes.treino <- ifelse(probabilidades.treino > 0.5, "Yes", "No")
> table(treino$Renuncia, predicted.classes.treino)
      predicted.classes.treino
      No      Yes
Não 1085    17
Sim   262    43
> sum(diag(prop.table(table(treino$Renuncia, predicted.classes.treino))))
[1] 0.8017058
> probabilidades.teste <- predict(rlog2, type = "response", teste)
> predicted.classes.teste <- ifelse(probabilidades.teste > 0.5, "Yes", "No")
> table(teste$Renuncia, predicted.classes.teste)
      predicted.classes.teste
      No      Yes
Não  463    12
Sim   114    14
> sum(diag(prop.table(table(teste$Renuncia, predicted.classes.teste))))
[1] 0.7910448
> ##### Avaliacao dos modelos construídos #####
> tab_cart <- ifelse(teste$Renuncia == "Sim",1,0)
> pred.cart <- predict(pcart, type = "prob", teste)
> confusionMatrix(tab_cart, pred.cart[,2], threshold = 0.5)
      0      1
0 469 116
1   6  12
> precision(tab_cart, pred.cart[,2], threshold = 0.5)
[1] 0.6666667
> sensitivity(tab_cart, pred.cart[,2], threshold = 0.5)
[1] 0.09375
> specificity(tab_cart, pred.cart[,2], threshold = 0.5)

```

```
[1] 0.9873684
> ks_stat(tab_cart, pred.cart[,2])
[1] 0.2135
> plotROC(tab_cart, pred.cart[,2])
> AUROC(tab_cart, pred.cart[,2])
[1] 0.6405428
> ks_plot(tab_cart, pred.cart[,2])
> plotROC(tab_cart, pred.cart[,2])
> tab_rlog <- ifelse(teste$Renuncia == "Sim",1,0)
> pred.rlog <- predict(rlog2, type = "terms", teste)
> confusionMatrix(tab_rlog, pred.rlog[,2], threshold = 0.5)
  0  1
0 425 111
1  50  17
> precision(tab_rlog, pred.rlog[,2], threshold = 0.5)
[1] 0.2537313
> sensitivity(tab_rlog, pred.cart[,2], threshold = 0.5)
[1] 0.09375
> specificity(tab_rlog, pred.cart[,2], threshold = 0.5)
[1] 0.9873684
> ks_stat(tab_rlog, pred.rlog[,2])
[1] 0.1368
> plotROC(tab_rlog, pred.rlog[,2])
> AUROC(tab_rlog, pred.rlog[,2])
[1] 0.5635444
> ks_plot(tab_rlog, pred.rlog[,2])
> roc.cart <- plot.roc(tab_cart, pred.cart[,2], main="Comparação de modelos", percent=TRUE,
xlim=c(100,0), col="blue")
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> roc.rlog <- lines.roc(tab_rlog, pred.rlog[,2], percent=TRUE, col="red")
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> legend("bottomright", legend = c("Árvore de decisão", "Regressão logística"),col = c("blue", "red"),lwd
= 2)
> roc.cart.ci <- plot.roc(tab_cart, pred.cart[,2], main = "Intervalo de confiança", percent=TRUE, ci =
TRUE, print.auc = TRUE)
Setting levels: control = 0, case = 1
Setting direction: controls < cases
```