****

INDEPENDENT REGULATORY AGENCIES (IRAs):

how consumers’ complaints relate to IRAs’ reputation?

Maiara Oliveira Marinho

Prof. Fabio Caldieraro

Machine Learning

2020

# *Introduction*

Since the 1990s, important debates among society about the regulation of several services led countries to increasingly seek to create more independent regulatory agencies (IRAs). During some decades, the key goals for countries was to create regulators that were really independent from political and industries’ interferences and that effectively reflect public interest (Wu, 2008). Independence, however, can be related to the regulators’ relationships with three main actors: the industry, the government and consumers. Our focus in this research proposal is the relationship between consumers and regulators, and how this relationship can affect agencies’ overall judgement, that is, their bureaucratic reputation.

Brazil has a strong reputation of independent regulatory agencies and is especially known for its efforts around the issue (Wu, 2008), but many agencies fail to acknowledge and to handle consumers’ long-term interests and we believe this can affect beliefs about them. Moreover, most studies do not focus on citizens’ judgements or regulatory decisions, but rather observe formal and organizational aspects of the regulatory bodies, like boards’ mandates (Peci, 2018), overlooking the importance of other conditions that also go around agencies’ effectiveness. For instance, an important research gap can be addressed by studies that investigate whether regulators are hold accountable for consumers welfare. Wu (2008) advocates that the only presence of special offices concerned with citizens’ interests inside IRAs can be considered an evidence of agencies’ alignment with the quality of consumers’ life, but we should be cautious about this indicator. Mainly because, based on the assumption that agencies are insulated from external influences, having an office directed to manage customers’ complaints may be somewhat a weak evidence of long-term orientation for consumers’ welfare.

This study also explores an important gap on literature which is measuring public organizations’ reputation using machine learning techniques. The importance of machine learning methods to Public Administration is not only becoming evident but is also becoming popular. There are already studies that address that using sampling frames like tweets (Anastasopoulos & Whitford, 2019), employing feedbacks from experts (Lee & Van Ryzin, 2019) and defining agencies’ priorities with data extracted from texts (Hollibaugh, 2019), but very few devote effort to looking at actual consumers’ assessment from a singular database like the one used in this proposal.

An interesting channel for receiving consumers’ complaints was established in Brazil in 2013, called consumidor.gov.br. It consists on a public service on the internet that allows direct dialogue between consumers and companies to solve conflicts. The tool is a website monitored by the National Consumer Secretariat (*Secretaria Nacional do Consumidor -* SENACON), from the Ministry of Justice, and by prosecutors, public defenders and other public bodies. In 2019, more than 780 thousand claims were handled, and 609 companies were registered on the website. Today, the number of companies exceeds 850. Officially launched in 2014, the platform has registered more than 2.5 million complaints from almost 1.8 million registered users. In the website, consumers can claim about any issue, like health, energy, communication and internet, and transport, all issues regulated by different IRAs.

According to SENACON, 80% of registered complaints at consumidor.gov.br are solved by companies, with an average response time of 7 days (Horttanainen, 2019). The complaints are addressed by many regulatory agencies, according to each competence, and SENACON establishes cooperation agreements with some of them. For instance, communication and internet problems are carried out by ANATEL, complaints related to food are addressed by ANVISA, problems with flights by ANAC, health insurance by ANS and so on. At the time of registering a complaint, the consumer fills in personal information and classifies the subject of the complaint by him(her)self using drop-down menus. At the end of the costumer service, the citizen gives scores to the solution provided, which will be our proxy for agencies’ reputation.

The contribution of this research and its empirical strategy relies on using this data of consumers’ claims from consumidor.gov.br in the last year (2019) to predict the scores citizens give to the solution provided and how they relate to agencies’ reputation. In other words, our focus will be predicting, using machine learning techniques (*Decision Tree and* *Random Forest*), the scores customers attribute to the solution companies and agencies provide to their problem, which range from 1 (poor solution) to 5 (excellent solution). The logic that supports this study comprehends that agencies are responsible for regulating and monitoring companies’ operation and, through cooperation agreements with SENACON, they are hold accountable for solved and unsolved complaints. Considering this, the scores citizens give to the solutions provided somewhat reflect agencies’ evaluation and, consequently, their reputation. By using machine learning techniques, we expect to know what the best predictors for clients are to attribute scores and relate them to each agency’s strong or poor reputation.

# *Background knowledge*

For more than two decades, Toyota was a company with a strong reputation, with cars that could last-long but after a recall for problems on the pedals of several popular vehicles, it faced a huge reputational damage (Lange et al., 2011). This story illustrates how organizational reputation is a quite intuitive concept, but with complex operationalization in Management studies. According to Langue et al (2011), three conceptualizations are important to understand organizations’ reputation: *being known,* which refers to the visibility of the firm; *being known for something,* which refers to the capacity of predictability of organizational outcomes; and *generalized favorability,* which refers to an overall judgement, that can be good, attractive or appropriate.

Reputation can also be understood as an intangible asset and strategic resource produced by the interaction with society over time in a complex social process (Deephouse, 2000; Lee & Whitford, 2012). It is centered on the evaluation of an organization’s character and activities by multiple audiences (Maor, 2015) and on how this evaluation may result in good general judgements (Lee & Whitford, 2012). Lange’s et al (2011) reading of the literature suggests that, although organizational reputation can be considered as an asset which is held by organizations, it is created and maintained in third parties’ cognitions.

However, some additional challenges emerge when we think about public organizations. Anastasopoulos and Whitford (2019) claim the organizational capacity to deal with these challenges depends on reputation because public agencies require support from political actors and need to project an image of “consistency and flexibility”. In the public realm, because “reputation forms a largely symbolic construct” (Carpenter, 2010, p. 45), it enables to public agencies fewer interventions (Lee & Whitford, 2012), grants power to enforce decisions (Carpenter, 2010) and connects the organization’s aspects of power, autonomy and legitimacy (Lee & Van Ryzin, 2019). According to Lee and Van Ryzin (2019, p. 178), reputation can also refer to overseer’s general perception of an agency’s “vigor and aggressiveness” when pursuing goals and operates by granting power to agencies to enforce their decisions less controversially.

To understand how reputation works, Carpenter and Krause (2012, p. 27) put it on the form of questions and each question is related to four critical-facets of the concept, that are: “Can the agency do the job?” relates to performative reputation; “Is the agency compassionate, flexible and honest?” relates to moral reputation; “Does the agency follow normally accepted rules?” relates to procedural reputation; and “Does the agency have the capacity and skill required for dealing in complex environments?” to technical reputation.

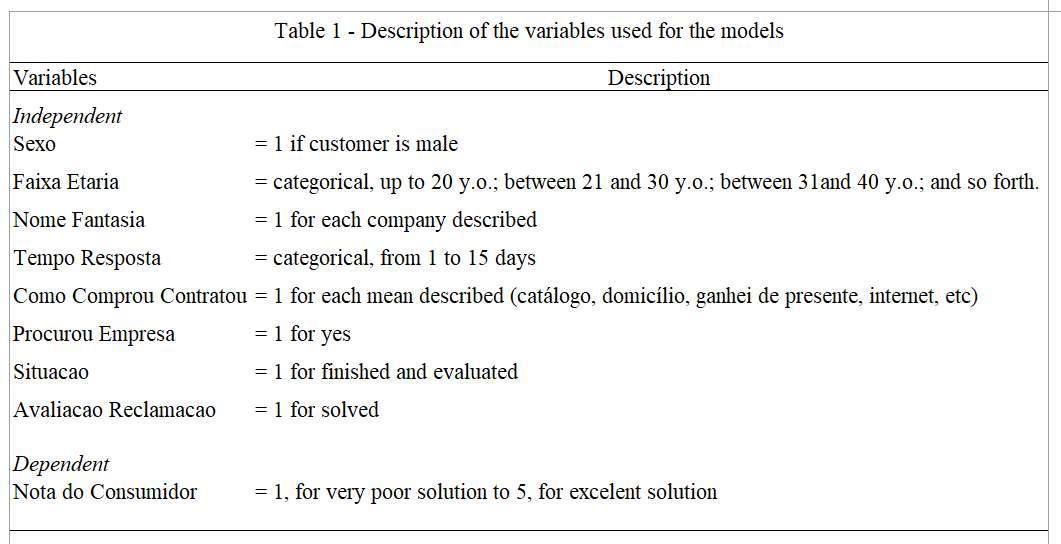
Some dimensions clearly overlap in some aspects, like moral and procedural dimensions, but Lee and Van Ryzin (2019) suggest that they may oppose to each other in a way that an agency can have moral ends, but procedurally deficient means; while it can act procedurally fair but with questionable ends. In its turn, technical reputation concerns an organization’s scientific authority, methodological sophistication and analytical capacity (Lee and Van Rizyn, 2019, p. 179), while performance reputation refers to the quality of decision-making to effectively achieve the agency’s goals, which may be or may be not considering consumers’ welfare.

In his *Reputation and Power: organizational image and pharmaceutical regulation at the FDA,* Carpenter analyzes the FDA (Food and Drug Administration) capacity to position itself as a public protector of national health using its reputation based on great expertise in a way that even elected officials could not challenge its authority (Miller, 2010). An important finding on Carpenter’s study was that, controlling by some epidemiological factors, he could evidence that drug review times were quicker when the disease treated by that drug was covered in the newspapers (Carpenter & Krause, 2012), suggesting that public criticism and emotional appeals also have a reputational cost for agencies. An important example occurred with the American president Bush and the “morning after pill”. Bush interfered strongly on the FDA and decreased its credibility because the pill was under attack by religious groups, perceived as an incentive for promiscuity (Miller, 2010). Another episode happened in the 1980s, when the FDA “lost” a battle for dying protesters with AIDS who confronted the FDAs procedures and tradition through media and professional organizations to change their drug reviewing divisions (Carpenter, 2010).

Nevertheless, in developing countries, regulation has exceeded its role of organizing industries, but was encouraged to answer universalization appeals and to ensure access to public services (Peci, 2018) and this assumption reinforces our idea of using data from customers to assess IRAs’ reputation. Our hypothesis is that consumers’ complaints can tell us a lot about agencies’ reputation, whether they are active, enforce normative acts and watch over consumers’ well-being.

# *Method*

This section covers the methodology that will be adopted to predict the scores citizens attribute to complaints at consumidor.gov.br and to understand how they relate to agencies’ reputation. First, we looked at the data and realized that some sectors, like health, had more non-solved complaints (with more than 4%) in 2019 than others. So, because most of the complaints (almost 40%) were concentrated on telecommunications, we chose to focus only on this area as a preliminary analysis of the data, and only then look at other agencies’ areas as suggestion for future studies. By intuition, the dataset about consumers’ complaints suggests looking to which features are better predictors for the scores and whether they relate to the regulation agency’s operation.

The data frame comprises complaints from the last years, since 2014, with more than 780.000 observations for each year and variables for region, state, gender, age, date, days for response, company, industry, sector, subject, problem, means by which he/she bought the product/service, whether he/she got in touch with the company, answered/unanswered claims, finished or not, and feedback scores, classified from 1 to 5, where 1 is for bad and 5 for good customer service. As previously mentioned, we used the scores as a proxy for agencies’ reputation because it is agencies’ duty to monitor and regulate the companies from their areas and SENACON maintains cooperation agreements with the agencies to monitor companies and ask them answers. All variables are categories that were self-classified by customers at the time of the complaint. The customers also detailed the claims in a box designed to describe the problem, elaborating and giving other elements with their own words. Because we are only interested on categorical and numerical data, we dropped the column “Problem”, where there the problem detailed description was and the information it provides was already identified in the columns “Area” and “Subject”. After cleaning the data, creating dummies for the variables of interest, our dataset ended up with 126.115 rows and 123 columns. The variables were described in the Table 1 below:

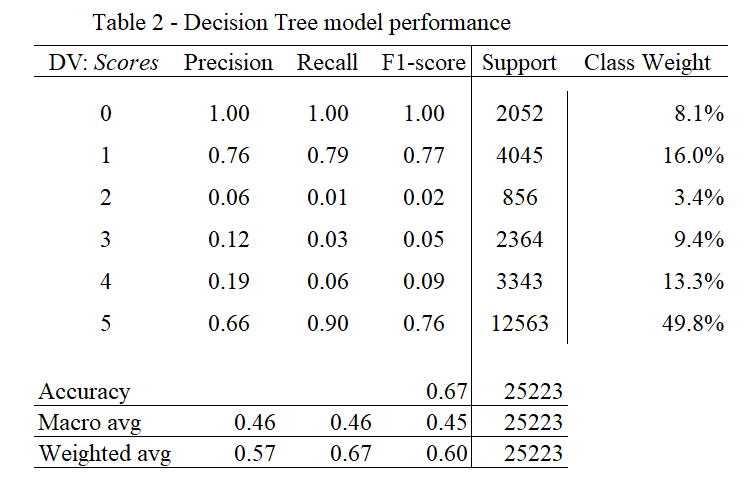
To assess and predict the importance of each feature, we chose two machine learning algorithms from Ensemble Methods: Decision Tree and Random Forest. According to Géron (2017), ensemble methods work best when predictors are independent from one another, but since the classifiers were trained on the same data, I used these two different algorithms to make the predictions.

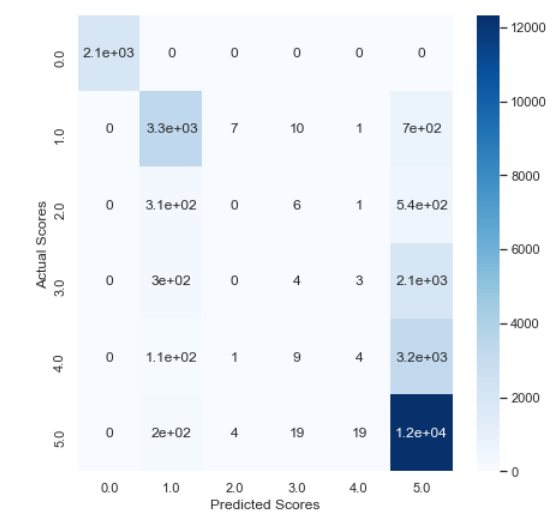
Decision trees use a tree structure to model the relationship between the features and are capable of finding complex nonlinear relationships in the data. Because we do not want to predict binary outcomes (the scores range from 1 to 5), we used a confusion matrix to report the performance of the algorithm. Random Forest, by its turn, is one of the most powerful Machine Learning algorithms and it does better than Ordinary Least Squares (OLS) because it searches for interactions automatically and introduces extra randomness when growing trees, that is, a random sample of *m* predictors is chosen each time a split is considered (Géron, 2017; James et al., 2013). Moreover, it searches for the best feature among a random subset of features, leading to a higher diversity and yielding an overall better model.

It is important to state that the column for our dependent variable contains zeros and null observations, but both mean that customers did not evaluate the service. After dropping the null values of “Nota do Consumidor” because many customers do not assess the claim after it is finished, our dataset ended up with 126.115 rows and 9 features. We expect that the decision tree and the random forest algorithms help determine which features are important to predict the scores of each complaint.

# *Using Decision Tree to predict the scores*

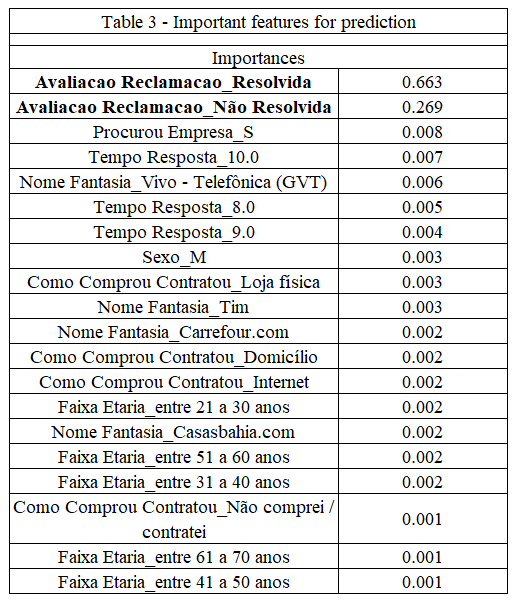
Since “Nota do Consumidor” is a numerical variable that can be treated as a discrete one, I directly estimated a decision tree model with the scores as the dependent variable, and sex, age, name of the company, time of response, how the consumption occurred, if the customer has looked for the company previously and the situation of the complaint, if it was finished or not, as independent ones. For the analysis, the ROC curve and AUC score are more suitable for binary classification problems but, as we are dealing with a multi classification problem, a more suitable metrics is the misclassification error. We chose the misclassification error because it represents the ratio of incorrect and correct predictions. Considering this, the best method will be the one that allows us to minimize the error. We also used a confusion matrix to present the results by the classifier.

For a maximum depth of 100, the classification report returned 33% of error and 66,8% of accuracy, as described below on Table 2. We could observe that the model accuracy rate is 0.66 and represents the ratio of correctly predicted observation to the total observations. The precision relates to the ratio of correctly predicted positive observations to the total positive observations. We can observe that it predicted 100% for 0 score; 76% for scores 1; 6% for score 2; 12% for score 3; 19% for score 4 and 66% for score 5 correctly. We believe that the model had low prediction rates for scores 2, 3 and 4 because our Train sample was an uneven class distribution and we could observe that on Table 2 on column “Class weight”. For the score 5, the model returned a precision of 0.66, which is above 0.5 and represents a good performance. The F1-score, the most useful measure, is the weighted average of precision and recall, which was good for the scores 0, 1 and 5 and we can observe this better prediction from the confusion matrix, in Figure 1, with darker blue squares.

*Figure 1 – Decision Tree Confusion matrix*

After that, we conducted an optimization using Grid search. According to Géron (2017), grid search is used to find the best hyperparameters using a cross-validation method. The parameters we used were: criterion (gini, entropy), max features (sqrt and log2) and max depth of the tree (10, 30, 50, 100). Using the measure of impurity entropy, Grid search returned the max depth of 10, and increased the accuracy to 70,23%.

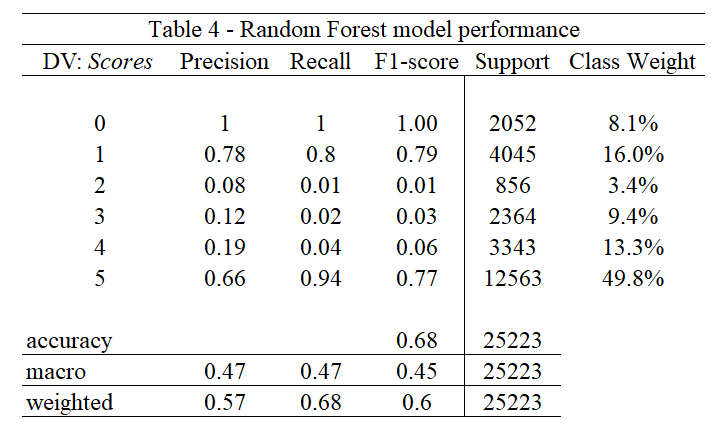
After assessing which features were most important for the prediction using Decision trees, we found an intuitive result: the variable “Avaliacao Reclamacao Resolvida”, with 66,3 percent, was the stronger antecedent for predicting whether the customer would assign a score to the service or not, and which score will be. In second place, we have got “Avaliacao Reclamacao Nao\_Resolvida”, with 26,9 percent. All the other variables had less than 1 percent of prediction power.



# *Using Random Forest to predict the scores*

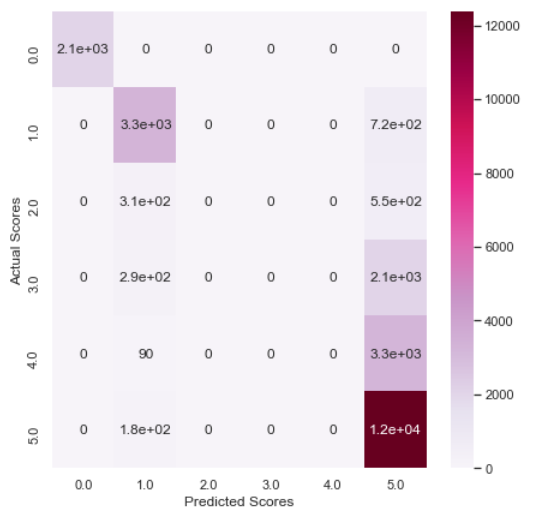
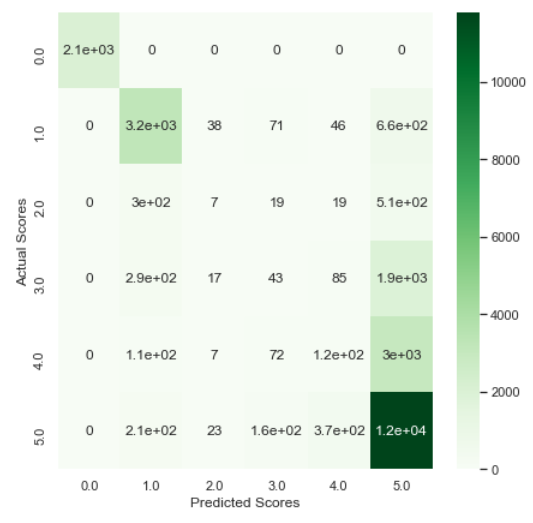
Using ensemble methods to predict the scores customers attribute to the service, to the solution provided, to the company and, consequently to the actuation of the regulatory agency, stems from the fact that multiple models give a better prediction than a single one. For both algorithms, we used a training set with 80% of the sample, with 100.892 rows, and 25.223 rows for the test set.

As previously indicated, it did not make sense to use the technique called the ROC curve to plot the true positive rates versus false positive rates because, according to Bowles (2015), the AUC gives a better overall measure of performance and our study is a problem of multiclassification. The ROC/AUC classifier picks three different threshold levels and the closer it reaches the upper-left corner, the best its prediction. If we chose the ROC classifier, it would return with True Positive and True Negative outcomes of interest, and we want to predict 5 classes together plus the zero score.

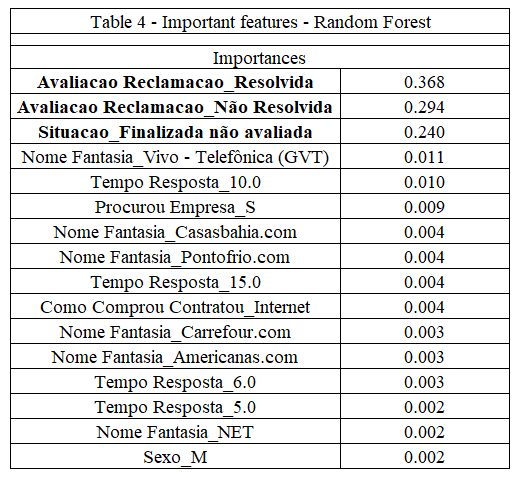
The Decision classifier only returns predictions of all individual trees, but such a Random Forest generates its sequence of models by training them on random subsets of data using the full training set (Bowles, 2015). Using 100 estimators, we wanted to predict the scores and we have got a classification report, which is represented on Table 4:

The Random Forest had a better hit rate, with 68,38%, than decision tree, but we could not see significant differences between both confusion matrixes. After optimizing using grid search also for the random forest algorithm, we have found a better hit rate of 70,86 percent, with the depth of the tree of 10 and number of estimators of 50. The accuracy scores increased to 70,43%. Moreover, we have got consecutively the importance of the features for the random forest model, and we could see some differences between the best predictors of decision tree. The features are described on Table 5 below. We can see that the most important feature using the Random Forest model still is “Avaliacao Reclamacao\_Resolvida”, which means that if the complaint was solved or not. However, it had a lower importance, with 36,8%, compared to the Decision Tree, with 66,3%. The second most important feature was still “Avaliacao Reclamacao\_Nao Resolvida”, which is the negative pole of the previous feature, with 29,4% of importance. However, a third feature appeared with high importance for the predictions, which was “Situacao Finalizada\_Nao Avaliada”, with 24%, that represents whether the complaint was actually finished, without considering if it was solved or not.

*Figure 2 – Random Forest confusion matrix Figure 3 – Optimized Random Forest*



*confusion matrix*



# *Discussion and Conclusion*

The present study aimed to predict the scores citizens attributed to the solutions provided on the consumidor.gov.br channel and relate them to the agency’s bureaucratic reputation using machine learning techniques, decision trees, random forest and gradient boosting. Considering the higher number of complaints related to telecommunications, with almost 37 percent of all the complaints in the year of 2019, we used the observations related to this issue. According to SENACON, the cooperation with ANATEL – National Telecommunications Agency has existed since 2013 and enables the implementation of joint actions aimed at protecting the rights of consumers of telecommunication services (ANATEL, 2017). In 2019, the average of scores given to telecommunication problems was 3,85, which is considered high. It is important to notice that 90,6 percent of the telecommunication complaints in 2019 was solved (SENACON, 2019) and it confirms the prediction of our models, which showed that an important predictor for high given scores was whether the claims were solved or not.

Based on our models and on the high scores observed on telecommunication complaints, we could infer that ANATEL has a strong reputation to enforce decisions to private companies and that agencies should watch over their norms and regulatory actions to get the complaints finished and solved. Another important aspect of the study was the possibility of using machine learning techniques to predict the scores, inputs provided directly by consumers, as an indicator of quality of services and strong reputation. By using a large amount of complex and real-world data, we could infer that ANATEL has a high standard of regulatory performance and reputation.

Therefore, a limitation of this study was using data only from 2019 and addressing only consumers’ complaints without considering the board composition of the agency and its bureaucratic, political and/or technical-regulatory orientation, which has already been shown to guide several decisions inside IRAs (Peci, 2018). Another essential feature of IRAs that was not considered in this proposal is the possibility of interests of organized groups capture the agencies ideologically, as already discussed by several authors (Carpenter & Moss, 2014).

Methodologically, this study used only two machine learning algorithms, decision that could be improved for future studies. In substantive terms, we would like to suggest addressing other agencies reputation based on the specific complaints directed at them, like aviation, health and energy, for example.

# *References*

Anastasopoulos, L. J., & Whitford, A. B. (2019). *Machine Learning for Public Administration Research, with Application to Organizational Reputation*. *29*(3), 491–510.

ANATEL. (2017). *Acordo de cooperação entre Anatel e Senacon é renovado*. https://www.anatel.gov.br/consumidor/tv-por-assinatura/direitos/fidelizacao/99-novidades/682-acordo-de-cooperacao-entre-anatel-e-senacon-e-renovado

Bowles, M. (2015). *Machine Learning in Python: Essential techniques for predictive analysis*. Wiley.

Carpenter, D. P. (2010). *Reputation and Power: Organizational image and pharmaceutical regulation at the FDA*. Princeton University Press.

Carpenter, Daniel P., & Krause, G. A. (2012). Reputation and Public Administration. *Public Administration Review*, *72*(1), 26–32.

Carpenter, Daniel P., & Moss, D. A. (Orgs.). (2014). *Preventing regulatory capture: Special interest influence and how to limit it*. Cambridge University Press.

Deephouse, D. L. (2000). Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories. *Journal of Management*, *26*(6), 22.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (1st edition). O’Reilly Media, Inc.

Hollibaugh, G. E. (2019). The Use of Text as Data Methods in Public Administration: A Review and an Application to Agency Priorities. *Journal of Public Administration Research and Theory*, *29*(3), 474–490.

Horttanainen, E. P. T. (2019). *Classificação automatizada de reclamações de usuários de serviços públicos: Um estudo do caso aéreo* [Masters Thesis] - IPEA.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York.

Lange, D., Lee, P. M., & Dai, Y. (2011). Organizational Reputation: A Review. *Journal of Management*, *37*(1), 153–184.

Lee, D., & Van Ryzin, G. G. (2019). Measuring bureaucratic reputation: Scale development and validation. *Governance*, *32*(1), 177–192.

Lee, S., & Whitford, A. B. (2012). *Assessing the effects of Organizational Resources on Public Agency Performance: Evidence from the US Federal Government*. *23*, 687–712.

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231.

Maor, M. (2015). Theorizing Bureaucratic Reputation. In A. Waeraas & M. Maor, *Organizational Reputation in the Public Sector* (p. 17–36). Routledge.

Miller, G. J. (2010). A Review of: “*Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA* by Daniel Carpenter”: Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*.* Daniel Carpenter. Princeton, NJ: Princeton University Press, 2010, 752 pages. *International Public Management Journal*, *13*(4), 471–475.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives, 31*(2), 87–106.

Peci, A. (2018). *Expertise e politização nas diretorias colegiadas das agências reguladoras brasileiras* (Escola Brasileira de Administração Pública e de Empresas, EBAPE/FGV). Policy Paper.

SENACON. (2019). *Boletim Consumidor.gov.br 2019*. https://consumidor.gov.br/pages/publicacao/externo/

Wu, I. (2008). Who Regulates Phones, Television, and the Internet? what makes a communications regulator independent and why it matters. *Perspectives on Politics*, *6*(4), 769–783.