XXII SemeAd
Seminários em Administração

novembro de 2019
ISSN 2177-3866

“Listen to Me”: Identifying Categories for Customer Complaint’s Mediation Automation

NATAL ACIR ROSA JUNIOR
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

CARLA BONATO MARCOLIN
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

VÉRICA MARCONI FREITAS DE PAULA
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

VERONICA ANGELICA FREITAS DE PAULA
UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

FERNANDA DA SILVA MOMO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Agradecimento à orgão de fomento:
We thank CNPq for the support for this work.

“Listen to Me”: Identifying Categories for Customer Complaint’s Mediation Automation

1. Introduction

The word-of-mouth posted online either on the website of the store itself or on other websites impacts consumer decision making (Koo, 2006), and represents a major difference mainly because of its long duration and long reach (Graham & Havlena, 2007). A relevant example of this phenomenon is the growing emergence of complaint sites, which concentrate on a single platform several comments and experiences of consumers from countless companies of the most diverse sectors, allowing the consumer a unique place for gathering information and opinions that help with the purchase decision. These platforms pose a challenge for companies as, according to Albrecht and Zemke (2002), the ability of organizations to handle customer complaints is a determining factor in the customer decision to choose the company again or to experience other offers. As a way of maintaining a synergy between customer and company and consequently being attentive to their relationship, it is important to have a complaints mediation department, performing a function that, according to Iasbeck (2012), can give the customer an appropriate attention by not only listening but also answering to its problems, difficulties and challenges.

Considering the importance of complaints in the healthy relationship between client and company, it is important to emphasize the importance of a department that knows how to receive and send to both sides the necessary information to solve the problem in a timely manner and avoid recurrence, which falls within the concept of ombudsman, defined as Braz and Varão (2012) as the sector that serves as an environment for receiving various information, capable of making known what would be imperceptible to the organization, as well as complaints, suggestions, praise and doubts. According to Demchenko et al. (2013), both the data and the possibilities of using them are constantly growing, in this sense, it is up to companies to use this data as a way to increase and improve their operation in order to improve processes and support decision making. In this case, an automated ombudsman model based on a processing, segmentation and classification tool can bring an effective gain of time and resources, making the customer service more efficient and agile, being able to serve more customers and possibly decrease the recurrence, as well as providing data and information that can improve the operation and results of the company as a whole.

This research aims to present the first step to develop a text analytic tool that enables the automation of a complaint mediation sector in a Brazilian retail company that sells home appliances and furniture. The first step, presented here, is the categories identification for the semi-supervised classification tool training. The company has over 180 stores in seven states of Brazil in addition to e-commerce. Through the use of computational software that allows the processing, classification and segmentation of 1,175 customer comments from the Brazilian website Reclame Aqui (“Complain Here”, one of the largest independent platforms regarding complaint sites in the country), the intention is to automatize the complaint sector by developing a semi-supervised classification tool that, for each new complaint posted in the mentioned website, automatically sends this message to the respective department, without the need of human mediation. Thus, we contributed, through this first step, to develop an environment where customers can have their problems solved more quickly and effectively. In addition, we bring to the academic literature evidence of how the process of automating data analysis can contribute to improvement customer service (Weitzl & Hutzinger, 2019), a challenge nowadays due to the number of customers, from different locations, with different needs.

2. Literature review

A public purchasing scenario makes manifestation in public or in front of third parties one of the ways to put consumer dissatisfaction in an important position regarding the process of buying or contracting services (Mowen & Minor, 2003). Considering the connectivity of the 21st century and the accessibility of consumers, there are some websites specialized in evaluations and complaints with great adherence of dissatisfied customers, which given the usability and effectiveness of the platform, concentrates a good amount of negative feedback information, allowing, for companies, a two-side effect being the negative exposition in front of other customers, as well as a good possibility of handling complaints and turn a negative into a positive opinion. Among the various specialized websites in Brazil, in particular, Reclame Aqui is one of the largest platform for public complaint register, as, according to data from the website (2018), "every day more than 600 thousand people research the reputations of companies before making a purchase, hire a service or solve a problem". The website is free of charge and allows that any registered consumer issue a complaint against calls, purchases, sales, products and services. Complaint that, in addition to being published, is sent to the company claimed (if it is registered), thus allowing an analysis and appropriate response to each situation.

It is relevant to mention that as the complaint platforms require a response, the company needs to be prepared to maintain a good synergy with customers, which, according to Iasbeck (2012), can be enabled by a complaint mediation department, that can shorten better handle customer dissatisfaction and provide answers and solutions. Braz and Varão (2012) emphasize that the performance of this department should be of an active conduct in order to be more effective in what concerns the communication of the society with the organization. According to Iasbeck (2012) the complaint mediation department is an integral part of the organization and should be responsible for an effective communication between the customer and the company, seeking to act preventively in order to avoid new conflicts. Iasbeck (2012) also points out that this department is the reality of almost all large and medium-sized companies in the Brazilian scenario, mainly because it is a fundamental part of the negotiation of obstacles in the customer-company relationship. In addition, this department is a valuable marketing tool that, according to Iasbeck (2012), as it seeks to ensure: consumer rights, supplier interests, increase customer loyalty and after-sales satisfaction, and serve as a kind of thermometer for preferences identification in consumption, operational problems, adequacy needs, among other things.

The complaints mediation department is then a strategic and fundamental sector for different companies, especially those in retailing. It obtains information that allows an organization to answer the questions from the external environment and it can also provide solutions and explain possible bottlenecks in the organization, providing suggestions for improvements and adaptations that enhance not only the operation but the results of the corporation. However, this department is only effective if it actually responds and solve customer problems, providing an adequate "webcare" (Weitzl & Hutzinger, 2019), what depends on getting the problem hands-on to the responsible department.

With large amounts of data available, it becomes challenging to be able to provide an adequate "webcare", once data volume and variety far outweigh the ability for manual analysis, while computers have become more powerful (Brown & Toze, 2017; Wu, Zhu, Wu, & Ding, 2014). The convergence of these phenomena gives rise, increasingly widespread, of data science principles and data mining in business (Fawcett & Provost, 2018). Inside this scenario there are several advances in handling unstructured data, specially text, giving companies the capacity to process and treat, automatically, high volumes of text data, the major format of user complaints.

Given data volume in complaint platforms, human mediation might be slow and somehow biased, what could interfere in this department value. By proposing an automated tool, there is an opportunity to contribute by providing not only the possibility to increase the volume of data analyzed, but also to overcome a step inside text analysis process, that implies a costly and quite subjective task of reading. Thus, it is possible to avoid bias that might bring different conclusions from the same data, once different people with different experiences might see different aspects when reading data in text (Ashton, Evangelopoulos & Prybutok, 2014).

3. Methodology

This research aims to present the first step to develop an automation tool for a customer complaint mediation department. In order to do so, we based on Design Science Research (DSR) (Hevner & Chatterjee, 2010), methodology, as it searches through data investigation, collected by standardized research techniques, to develop a tool that can help a specific problem in a business context. In addition, it is also characterized by the use of the concepts of quantitative analysis, which, according to Richardson (1999), counts on the use of quantification both in the information collection and in their treatment using statistical and mathematical techniques.

According to the checklist proposed by Hevner and Chatterjee (2010), a first issue to be clarified was the organization's problem. It was found that the current operation of the customer service receives connections of the most diverse characteristics, but usually is not able to solve and / or mediate the relevant issues of the operation, being restricted basically to answering the calls. It does not delegate the problems to the areas or departments that could solve them. In this case, the tool to be developed in this research could in fact add value to practice and potentially help managers understand the main problems reported by the customers.

Thus, a first step was the data collection, carried out in May 2018, where 1,175 comments were extracted from the Reclame Aqui website, being that data complaints registered from June 2014 to October 2017, provided by the company studied. The data was received a spreadsheet, containing all the fields taken from the website Reclame Aqui. The dataset contained 30 columns, that for this research purpose, was reduced to three columns: Date of the Complaint; Title and Text of the Complaint, being those last to concatenated to a single text field. Table 1 presents an original data sample, translated to English.

Table 1: Data Sample

Date	Title	Complaint text
June 8th, 2017	Delay in the delivery	I bought in the storeabc an Electric Oven Crystal Plus Layr because I saw that it was cheaper but I do not know how a store so famous can advertise a product for the customer to buy on the site being that it is not available to prompt delivery. Not the first time I buy products online, I know that when the product is missing, the virtual store warns us that the product is UNAVAILABLE FOR PURCHASE. But it was available for purchase […]
June 8th, 2017	Product was not delivered on schedule	I bought it in the store and the deadline for the delivery has already expired and until today they did not deliver it, nor gave any explanation, I do not recommend anyone to buy in this store.

Date	Title	Complaint text
May 24th, 2017	I NEED TO SPEAK WITH SOMEONE FROM THE STORE	I bought a product, I have some doubts and I did not find any 800-numbers to help me, if someone can contact me I would appreciate it very much

Source: Research Data

After the data gathering, a pre-processing was performed, in order to clean the data, by removing special marks and punctuation; capital letters; blank spaces and stopwords, as suggested by literature (Aggarwal & Zhai, 2012). This last procedure consists of removing very frequent words that do not bring relevant information to the construction of the tool, such as: "a", "of", "the", "that", "and" among others. Also, grammatical errors, among other characters that could disrupt the classification of the data, were removed. After performing these processes, the data could finally be transformed into a matrix to be manipulated.

In order to help with the organization´s problem, the tool will be structured to automatically classify the complaint in a certain category. For the identification of the categories, an analysis of the comments was sought so that they emerge from the text itself, helping not only define more effective categories, but also helping the organization to understand its main points of improvement. For this process, Lexicon Analysis and Latent Semantic Analysis was performed in order to better understand the structure of the data.

Among the several options of tools directed to statistical analyzes such as R, Tagul, Python, etc., the chosen option was the RStudio software, a tool of the R system package, of development of decision support systems and data analysis. This software is available in open source and allows the user to modify it to the functionalities that are needed in the applied research. Besides being a very effective tool, it includes in its package several integrated development environments, such as R-Commander, R-Shell, R-Python, and others. All of those are designed for statistical and econometric functions.

In addition, we performed an investigation using Latent Semantic Analysis (LSA). LSA is an unsupervised techniques part of a group of methods called Topic Modeling, that aim to uncover the expression patterns in each category. In order to connect ideas (or topics) besides differences between words, LSA uses singular-value decomposition (SVD). SVD is a decomposition solution to deal with non-square matrices. This decomposition is based on vector space models where term weights and the representation of documents as vectors in a space are possible, which allow the application of concepts such as measures, distances and similarities between documents and words (Wild, 2016; Kulkarni *et al.*, 2014). With that mathematical structure, LSA allows to understand the most important topics in a set of documents, which is suitable with our objective, that is, to emerge from the text the set of categories connected with the most common problems that customers complain.

Knowing the categories and its main words is fundamentally to have a consistent classification. After the categories were established, we had three manual classification rounds with three different specialists (two from the company and one researcher) using an aleatory sample of 100 comments in order to develop a classification protocol that was then used to classify all 1,175 comments. After that, we apply machine learning techniques in order to train and test classifiers.

For this task we applied the machine learning algorithm known as Support Vector Machine (SVM) (Vapnik, 1995) that, based on the concept of decision plans, aims to define a limit which separates objects from different classes (Vidhya, 2013). This method has obtained several positive results in classification problems such as handwriting recognition, face and image detection, text categorization, among others (Lima, 2014).

The SVM is a statistical learning machine that implements the Structural Risc Minimization (SRM) principle, delimiting an optimal hyperplane that maximizes the margin of separation between data classes in linear data problems. Due to its good performance, the SVM has become an excellent research reference in Machine Learning (Soliman, Mahmoud, 2012).

One of the guidelines pointed by Hevner and Chatterjee (2010) is the necessity to provide verifiable contributions. Thus, in order to improve the practical application, an interview was conducted with the manager responsible for the customer service, using an executive resume containing the main results from the research, which allowed a greater understanding about the possible applications in the daily operations.

4. Results

After data pre-processing, the first analysis consisted of understanding the lexicon of the database. Being a polarized base, once it has been collected from a complaint's website, a negative reaction is common throughout all the data file, mainly because the data consist of a set of strictly negative feelings, mostly with complaints. Considering this fact, the word "no" was removed from the database in order not to bias the study, thus providing a more efficient analysis. In addition, in order to preserve the company, its name was changed to storeabc.

Throughout the database, the words that stood out the most were "Day," "Product," "Delivery," and "Order". Together, these expressions were responsible for over 7% of the total frequency. Considering a total of 8,389 words, if all words had the same weight, four words would count for last than 0.04%. It is reasonable to say that these expressions demonstrate the most prominent problems in the analyzed complaints, especially the delay of delivery.

Table 2: Ranking of the 20 most cited words

#	Word	Frequency	#	Word	Frequency
1	Day	1586	11	Storeabc	699
2	Product	1526	12	Act	677
3	Delivery	1152	13	Contact	600
4	Order	975	14	Bought	578
5	Days	934	15	Useful	497
6	Purchase	918	16	Because	480
7	Company	785	17	Received	451
8	Website	784	18	Email	443
9	Deadline	775	19	Payment	391
10	Store	736	20	Nothing	391

Source: Research Data

In order to complement this analysis, we searched for the words most connected with those top-4 by calculating the correlation between them given a term-document frequency matrix. Table 3 presents the top-10 words most correlated with each of them, as well as the correlation index. With that, it is possible to better understand the context in which those words were used. The first one, "Day", is more related with delivering aspects such as deadline, until, received and home – this late probably a reference to be or not to be at home during a specific delivery. In the second, "Product", we notice some of the usual steps when something is wrong with the product received, such as taking and sending photos or a box with some problem, like a strike. Next, "Delivery", is highly correlated with deadlines, in a sense that appears to be close to the ideas behind those words connected with “Day”. Finally, “Order” brings aspects given the purchase process, such as order cancelation, credit card issues and status consultation.

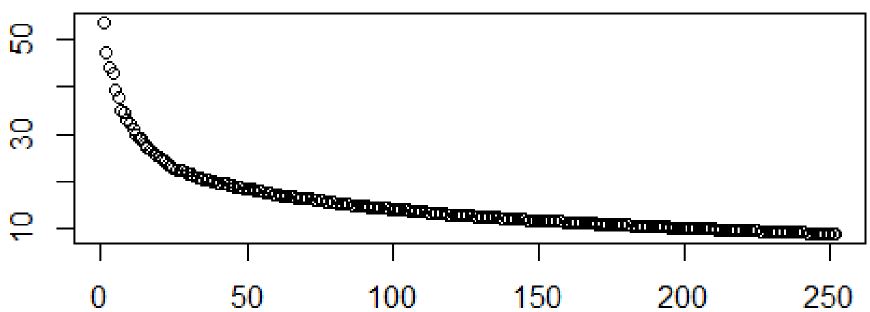
Table 3: Top-10 words correlated with top-4 raking words

Day		Product		Delivery		Order	
today	0.35	photos	0.30	deadline	0.38	canceled	0.31
deliver	0.33	send	0.28	Due	0.36	number	0.25
deadline	0.32	box	0.26	Day	0.33	status	0.25
until	0.30	receive	0.25	days	0.31	card	0.23
email	0.26	talked	0.25	shipping co.	0.24	was	0.23
payment	0.25	days	0.24	Late	0.22	sent	0.21
received	0.25	exchange	0.23	delivered	0.22	approved	0.20
call	0.25	taken	0.23	confirmation	0.21	cancellation	0.20
home	0.24	delivered	0.22	product	0.2	store	0.19
protocol	0.24	store	0.21	home	0.20	website	0.18

Source: Data Research

Finally, we performed the LSA analysis. First, from the singular values plot, we noticed that less than 50 from the 253 dimensions were responsible for most of the data variability, which means that there is a high concentration of information in few categories – which is expected when working with text data (Zipf, 1949). We choose to analyze the first 25 dimensions, looking for what subjects or categories they represented.

Figure 1: Singular Values. The x-axis is the number of singular values (total of 253), and the y-axis indicate its raw value. The higher the value, the more this dimension represent in total data variability.



Source: Research Data

The most relevant conclusion was that in all first-25 dimensions, there was the presence of at least one of the top-4 words mentioned above ("Day," "Product," "Delivery," and "Order"). Not only that, but those words were among the top-10 in each category, what illustrates that these words are significantly meaningful given the topics in the database. Table 4 presents the top-5 words from the top-10 left eigenvectors, which represents the words most connected with the 10 highest singular values.

Table 4: Top-5 words from top-10 singular values (SV)

SV 1	SV 2	SV 3	SV 4	SV 5	SV 6	SV 7	SV 8	SV 9	SV 10
Product	product	store	store	order	days	contact	order	contact	due
Store	days	day	Day	website	money	until	contact	purchase	storeabc
change	deadline	delivery	purchase	delivery	order	storeabc	card	days	order
purchase	day	due	card	purchase	email	fedex	inside	inside	nothing
photos	delivery	deadline	storeabc	number	due	order	product	due	today

Source: Research Data

With all analysis compiled, we concluded that those four words were representative of the main problems faced by this store. With that, those should be categories in which complaints could be automatically designated, in order to reach faster the correspondent department. Considering the actual structure of the store and the data, we decided to compress the "Day" and "Delivery" aspects in a single category, "Delivery", since we understand that those were problems related with delivery deadlines, a major issue considering the database. The other two categories were "Product", related with product defect, and "Order", compressing purchase and website issues.

For the work of the SVM algorithm, we used the proportion of 80% of the 1,175 comments available for training and the rest for classification, i.e., 940 for training and 235 for classification. Using this criterion, we used 15 different seeds (set.seed) provided by indications from third parties not involved in the process (seeking a greater randomness of numbers, not biasing the selection of seeds). The different seeds were used to randomize the data, making that in each turn different documents would be in different groups (training and test).

Since that we were dealing with high sparsity data, as text data nature, we decided to develop three classifiers, one for each category (Delivery, Product and Order). In this structure all 1,175 comments were classified with 0, when not belonging to the category, and 1, when belonging to the category, resulting in three different data entries for three classifiers.

At the end of each classification round, a confusion matrix was generated in order to evaluate the classification results. A confusion matrix (Manning *et al.*, 2009) shows true positives, that are correct classification of the class positive (belonged to the class to which it was classified); false negatives, which are errors in the model, that classified in the negative class when the expected value was class positive (it was classified as not belonging to the class when it should be owned); and finally false positives, that are errors in which the model predicted the class positive when the expected value was negative class (where it was classified in a class to which it did not belong); as demonstrated in table 5.

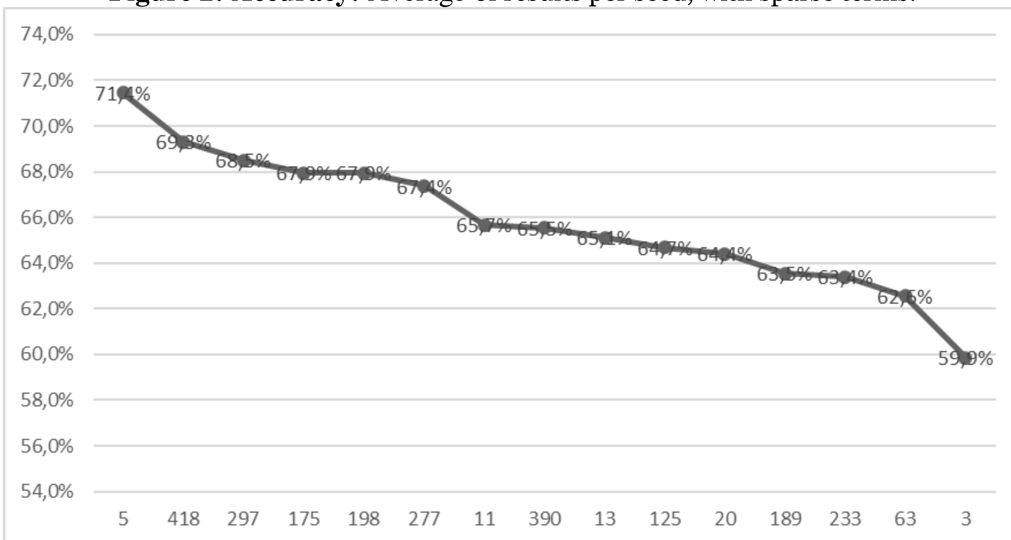
Table 5: Confusion matrix

	Class Zero	Class One
Class Zero	true positive (tp)	false positives (fp)
Class One	false negative (fn)	true positive (tp)

Source: Adapted from Manning, Raghavan and Schütze (2009), p.155.

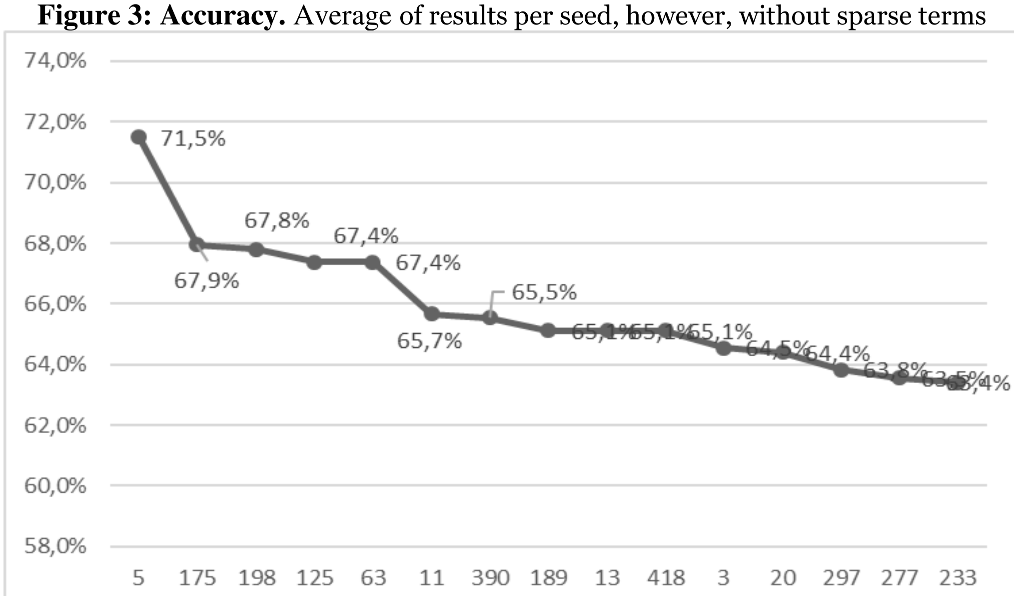
In order to measure the model performance, we measure the accuracy of the model. The accuracy represents the fraction of the classifications that are correct (Manning, *et al.*, 2009), which can be mathematically described as the proportion of true positives regarding all data (true positives + false positives + false negatives). Additionally, we tested the performance difference when considering or not considering sparse terms, i.e., those terms that represented less than 1% of all terms in all 1,175 documents. The results are in Figures 2 and 3.

Figure 2: Accuracy: Average of results per seed, with sparse terms.



Source: Research Data

Figure 3: Accuracy. Average of results per seed, however, without sparse terms



Source: Research Data

With an accuracy rate raging from approximately 72% to 60%, the classifiers developed were able to perform better when removing the sparse terms. As shown in the, the best result was found with seed 5 in the classification made after the removal of the sparse terms, where the mean of hits was 71.5%, which followed the evolution, considering that the average of the hits with the sparse terms were smaller than those defined after removal (65.8% with sparse terms versus 65.9% without).

With the research results, we elaborated an executive resume and presented to the manager responsible for the customer service. The first comment was related with the accuracy rate, that was considerable better than the effectiveness that the sector has today, with five employees:

> "In theory, the ombudsman would be responsible for internal service, being aware of complaints from the employees and the Operations Department, while the SAC, which responds to the Commercial Directorate, would be restricted only to customer service. In practice, both sectors receive calls from all publics, and in general, although both respond to requests and receive complaints, none can resolve the situation or have a regular route to sectors that could be resolved".

The conclusion was that the classifiers would allow better decision-making process, transforming a point of conflict in an opportunity to gain positive evaluation for the company. When asked about process that could gain with the model implementation, the manager saw opportunities in indexes generation, being able to evaluate the complaints and resolution by a dashboard. Finally, when asked about challenges to implement the solution, he saw none, mentioning that "quite the contrary, all techniques are not only valid but necessary as far as practice is concerned".

5. Final Considerations

With the large volume of available data related to customer printing on the various variables that involve a purchase, it becomes increasingly important to have a department that knows how to receive and send the information needed to solve the problem on both sides. in a timely manner and avoid recurrence (Braz and Varão, 2012). In this scenario, this research aims to present the first step to develop a text analytic tool that enables the automation of a complaint mediation sector in a Brazilian retail company that sells home appliances and furniture.

Through the use of computational software that allows the processing, classification and segmentation of 1,175 customer comments from the Brazilian website Reclame Aqui ("Complain Here", one of the largest independent platforms regarding complaint sites in the country), we develop the first step to automatize the complaint sector by developing a semi-supervised classification tool. We identified that most of the categories of analysis were related to delay of delivery. With LSA analysis, we conclude that in all first-25 dimensions, there was the presence of at least one of the top-4 words ("Day," "Product," "Delivery," and "Order").

With all analysis compiled, we concluded that those four words were representative of the three main problems reported by the customers. With that, those should be categories in which complaints could be automatically designated, in order to reach faster the correspondent department, contributing to develop an environment where customers can have their problems solved more quickly and effectively. So, with these results, we bring to the academic literature evidence of how the process of automating data analysis can contribute to improve customer service (Weitzl & Hutzinger, 2019), once knowing the categories and its main words is essential to have a consistent classification and, thus, improve the relationship with clients.

Data was restricted to a single company, and from a specific period, which can be seen as research limitations. In addition, other complaints received directed from the company's SAC were not considerable and could have brought other categories to analyze.

Future research, thus, plans to combine external and internal data in a single classifier, in order to expand the capacity to map main category complaints and deliver to the adequate department to solve the problem as fast as possible. In addition, we can mention the reassessment of the algorithm through the wrong classified percentage, and search for other artifacts that search for a better percentage of accuracy, whether for new computational software such as Python or other algorithms such as Naïve Bayes and Neural Networks, given its reported performance with text data (Aggarwal & Zhai, 2012) in order to compare performances given data structure.

REFERENCES

Albrecht, K., & Zemke, R. (2002) Serviço ao cliente: a reinvenção da gestão do atendimento ao cliente. Rio de Janeiro: Campus.

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.

Ashton, T., Evangelopoulos, N., & Prybutok, V. (2014). Extending monitoring methods to textual data: a research agenda. Quality & Quantity, 48(4), 2277-2294.

Braz, A., & Varão, R. (2012) O papel das ouvidorias na comunicação organizacional. In: Ouvidoria: mídia Organizacional. Porto Alegre: Sulina.

Brown, D. C. G., & Toze, S. (2017). Information governance in digitized public administration. *Canadian Public Administration-Administration Publique Du Canada*, *60*(4, SI), 581–604.

Demchenko, Y et al. (2013) Addressing Big Data issues in scientific data infrastructure. First International Symposium on Big Data and Data Analytics in Collaboration (BDDAC 2013). Part of The 2013 International Conference on Collaboration Technologies and Systems (CTS 2013), May 20-24, 2013, San Diego, California, USA.

Fawcett, T., & Provost, F. (2018). Data Science para negócios: o que você precisa saber sobre mineração de dados e pensamento analítico de dados. Rio de Janeiro: Alta Books Editora.

Graham, J., & Havlena, W. (2007) Finding the “Missing Link” Advertising’s Impact on Word of Mouth, Web Searches, and Site Visits. Journal of Advertising Research, 47(4), 427-435.

Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In Design research in information systems (9-22). Boston: Springer.

Iasbeck, L. C. A. (2012). Ouvidoria, Lugar Privilegiado de Comunicação Organizacional. XXIX Congresso Brasileiro de Ciências da Comunicação.

Koo, D. (2006). The Fundamental reasons of e-consumers’ loyalty to an online store. Electronic Commerce Research and Applications, 5(2), 117-130.

Kulkarni, S. S., Apte, U. M., & Evangelopoulos, N. E. (2014). The use of latent semantic analysis in operations management research. Decision Sciences, 45(5), 971-994.

Manning, C., Raghavan, P. and Schütze, H. (2009). Introduction to information retrieval (154-163.) Cambridge: Cambridge University Press.

Mowen, J. C., & Minor, M. S. (2003). Comportamento do consumidor. São Paulo: Pearson Prentice Hall.

Reclame Aqui. (2018). Institucional. Retrieved from: https://www.reclameaqui.com.br/institucional/.

Vapnik, V. (1995). The nature of statistical learning theory. Berlin: Springer-Verlag.

Vidhya, M. (2013). Efficient classification of portscan attacks using support vector machine. In Green High-Performance Computing (ICGHPC), 2013 IEEE International Conference.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Oxford, England: Addison-Wesley Pres.

Weitzl, W. J., & Hutzinger, C. (2019). Rise and fall of complainants' desires: the role of pre-failure brand commitment and online service recovery satisfaction. Computers in Human Behavior, 97, 116-129.

Wild, F. (2016). Learning analytics in R with SNA, LSA, and MPIA. Springer.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107.