Analysis of the feelings of the population's opinion in social media: a look at education

Viviani Kwecko
IFRS- Rio Grande Campus
Federal Institute of Rio Grande do Sul
Rio Grande, Brazil
viviani.kwecko@riogrande.ifrs.edu.br

Fernando P. de Tôledo
Computer Science Center
Federal Univerity of Rio Grande
Rio Grande, Brazil
ftoledo@furg.br

Sam Devincenzi
Computer Science Center
Federal Univerity of Pampa
Alegrete, Brazil
sam.devincenzi@gmail.com

José O. de S. Ortiz
Computer Science Center
Federal Univerity of Rio Grande
Rio Grande, Brazil
joseoxlei@gmail.com

Silvia S. da C. Botelho
Computer Science Center
Federal Univerity of Rio Grande
Rio Grande, Brazil
silviacb@furg.br

***Abstract*—This research presents a work in which we identify and systematize how the vertiginous growth of social media allows the monitoring of public opinions, with a special focus on analyzing the feelings of the population's opinionated arguments about Education. We have brought together different methods in order to produce better results for the classification and summarization of various documents considering education as the basis of analysis. The proposed model is based on the steps of i) classification of patterns based on Deep Learning; ii) analysis of contexts and visualization of different associative paths in publications through the Implicative Statistical Analysis; and iii) validation of opinion abstracts. The results presented in this study refer to the database made up of 42,062 publications related to the city. The collective social discourses, resulting from the analysis of the summarize the opinions of 820 posts that presented representative terms for the education axis in the negative polarity, of the total of 975 posts classified by the dataset.**

***Index Terms*—Sentiment analysis, Text summarization, Public Opinions, Education Monitoring**

I. INTRODUCTION

This research presents a work in which we identified and systematized how the vertiginous growth of social media allows the monitoring of public opinions, with a special focus on the analysis of the population's opinionated arguments on Education, displayed on social networks. A national survey on the opinion and evaluation of the population regarding public policies and services for education, was developed by the Institute for Applied Economic Research and organized using a System of Social Perception Indicators (Sips[1]). This study traced a profile of the population's conceptions using 21 questions applied in the residence to only 2,773 people throughout the Brazilian territory, with the Brazilian population at the time being 190 million inhabitants (Censo 2010[2]). A decade later, we noticed a lack of continuity in this evaluation, which generates a lack of knowledge on the part of the public administration about the real meaning of Education for the population.

[1], [2] and [3], emphasize that the importance of an opinionated analysis of the population is justified by the support that this knowledge can give to teachers and administrators. Especially when discussing problems related to school structure, development of curriculum content, instructional methods, organization and administration of education. Furthermore, the approximation between the view of public opinion and that established by the school community about the educational objective to be achieved, may favor an increase in social support for the improvement of education [3].

Currently, we realize that Information and Communication Technologies (ICTs) have generated a large number of data capable of offering a profile of the massive perception of the population in real time [4], [5]. Access to this huge amount of opinions produced and published on the web, poses the challenge of a structural validity for such content. Identifying, monitoring and applying a large volume of opinionated text is a complex task [6], [7] due to: (i) the multiplicity of sites on which these "lines" are expressed; (ii) for the difficulty in identifying the relevant themes of the opinions, and (iii) the diversity of the information contained in the multiple comments.

In appendage to these structural obstacles, it is necessary to consider that all information is subject to interpretive biases that are most often consistent with the preferences of the data analyst [8]. Bringing together a set of Mining techniques [9], Opinion Classification [10], and Text Summarization [11], automated evaluation systems come to represent potential tools for understanding opinionated content. Mainly, by offering strategies for structuring an organization of the content of opinions in the midst of social networks, as well as seeking to overcome some of the subjective biases present in human analysis of such [8] content.

[1]http://www.ipea.gov.br/portal/index.php?option=com$_{c}ontentview = articleid = 12660catid = 4Itemid = 2$

[2]https://censo2010.ibge.gov.br

In this sense, in order to extract, structure, understand the information we approach some computational and statistical approaches, to quote: i) Data mining [12]; ii) Text mining [13], the use of Deep Learning Techniques for natural language processing (from English, Natural Language Processing) applied to classify [14], extracting context [15] iii) Implicative Statistical Analysis (A.S.I.), a statistical method of multidimensional data analysis in Education [16].

In section 2 we present the methodological approaches of the study for the analysis and production of collective discourses. In section 3 we present the methodology adopted for the case study of the speeches resulting from the application of the different approaches. In section 4 we present the results obtained. Finally, in section 5 we present the conclusions and discuss perspectives for future work.

II. Summary of Opinion as a contribution to decision in Educational Contexts

In contemporary cities, citizens' interests and demands emerge from different channels available for issuing their opinion, such as e-mails, ombudsmen, complaint portals and specific social media groups [17]. For [18], discovering public opinion is a significant factor in guiding governance actions in relation to citizen participation. However, it is important to differentiate public opinion that brings together a collection of ideas and values from a social representation that brings together an organized and structured set of information, beliefs, opinions and attitudes.

[19] warns that public opinion is not easily susceptible to scientific definition, since it represents a by-product of educational processes, as well as the growth of the mass media. [20], in turn, understands the concept as a process completely different from the public framework. For the author, opinion is a momentary and, in a way, logical grouping about judgments that start to be reproduced in a given circumstance, in a certain time-space and restricted to a social nucleus, representing the passage from an individual opinion to a collective opinion. [21] also seeks to conceptualize and theorize public opinion, coming to the conclusion that public opinion concerns the whole that we share with each other involuntarily and where we can intervene more and that originates a collective opinion. Another researcher who tries to define public opinion, who determines that opinion at the individual level is an attitude and at the collective level a feeling of the people [22].

Besides that, we found a vast literature in the field of public opinion analysis [23]; [24], related to the study of customer / user satisfaction [25], aimed at qualifying social- policies [26], [27], monitoring and improve education [28], [29] , governance [30] and, with greater intensity, aimed at market and product research [31]. However, since the works of [2], we have identified few attempts to systematically study the opinions of the lay public in relation to controversial issues, especially those that affect the field of education. Industrial organizations are interested in obtaining an ongoing measure of consumer opinion in relation to their specific product; government departments and agencies seek to obtain public opinion in relation to their administrative policies, so that they are able to pre-test public reaction before a policy or program is initiated; and executives try to get complete information about the public's current attitudes and then target educational advertising to succeed in their goals.

In view of these predilections, we question how the opinion poll of the lay public can be applied to community participation tools in the implementation of public policies for Education. Teachers and administrators are alert to public opinion. The usual practice is to listen more closely to individuals and organizations that are important to the educational community [3]. However, this listening is rarely close to the disorganized opinion expressed by the majority of the community. Often, the opinions of a well-organized minority are able to bring about educational changes simply because they are militants of a program they want to reach [1], [2]. Several political actors represent the public in the process of forming policies, according to [32], in a more general sense, we treat the public as an amorphous and passive entity, measured by attitudinal research that serve in the political process as pressure mechanisms for certain objectives.

In this study, the challenge is to understand the points of convergence and divergence between a set of opinions published on digital networks and their ability to reveal collective intelligences for the management of public policies in Education. The production of the text summarization model generated by the machine proposed in this research considered in its development the analysis of data related to the Education axis. We adopted as a summary the act or effect of extracting and synthesizing information from a set of texts [33]. Our work differs by combining in its stages different phases of the summarization typologies, through which we seek to identify the content associated with social intelligence attributes applied to education management.

In this context, opinion analysis generates information that can be used by various actors, whether they are representatives of the government or of social interests. For education, [3] identifies the potential for the use of public opinion measurement techniques as a tool to be used by teachers and administrators in determining problems related to school support, curriculum content, instructional methods, organization and administration of education. According to the author, when school administrators and teachers began to use community surveys to determine basic aspects of school management.

III. Data Mining for Decision Support

Every day a huge amount of data has been generated by devices, social networks and several other software applications and digital devices. This massive generation of data results in "big data". The analysis of these data has provided companies and individuals with interesting opportunities [34].

In [35] the fact that companies like *Google*[3] and *Microsoft*[4] use big data for business analysis and decisions, impacting

[3]https://www.google.com.br/

[4]https: //www.microsoft.com

existing and future technology. The researchers [5] highlight, in turn, studies that use big data in data coming from social networks. The use of Big Data in content analysis from social networking sites opens up the monitoring of opinion for what [36] considered as ”new era”. As a convenient source of user opinions, interactions and behaviors, networks expand the ability to examine social data on a large scale and in short periods of time.

These characteristics have aroused the interest of many researchers who seek to understand, for example, social relationships and behavior [36]–[40]; large-scale contagion processes [41], [42]; tracking preferences and / or large audiences [43]; social behaviors and attitudes [44]; collective experiences based on a timely event [45], [46]; the collection of large amounts of data on hard-to-reach populations [47]; mapping mood swings and other feelings [39], [48]; and research that points to the potential of social networks for the production of intelligence in the city is linked to the organization of social movements and the internet [49].

Identifying, monitoring and allocating a large volume of opinionated text is a complex task [6] due to: (i) the multiplicity of sites on which these ”lines” are expressed; (ii) for the difficulty in identifying the relevant themes of the opinions, and (iii) the diversity of the information contained in the multiple comments. In addition to these structural obstacles, it is necessary to consider that all human analysis of information is subject to interpretive biases that most of the time are consistent with the preferences of the data analyst. Automated systems of classification, opinion mining and text summarization represent, therefore, potential tools for the construction of a process of understanding opinionated content from social networks, allowing the structuring of data and overcoming some of the subjective biases.

Thus, the focus of this research falls on the need to develop monitoring mechanisms for the opinions published in groups of sites that discuss the city’s educational problems so that the information produced can establish indicators for public management planning. Therefore, it is essential to structure a process of collective citizen listening to identify, classify and assign the themes present in each opinionated post and to associate the contents in social representations of contexts.

IV. Methodology

This work assumes the case study as a research strategy whose methodological conception aims at a quanti-qualitative analysis [50] from the exploratory typology. We adopted a model, figure 1, for extracting and organizing opinion-oriented information [8]. The model brings together different methods seeking to produce better results for the classification and summarization of various documents related to the Education axis. The proposed model is based on three stages: *i) classification of patterns based on Deep Learning techniques; ii) analysis of the representation of contexts and visualization of different associative paths; and iii) validation of opinion abstracts*.

The classification step is taken from the application of a C-LSTM deep neural network model [51]. Used in conjunction

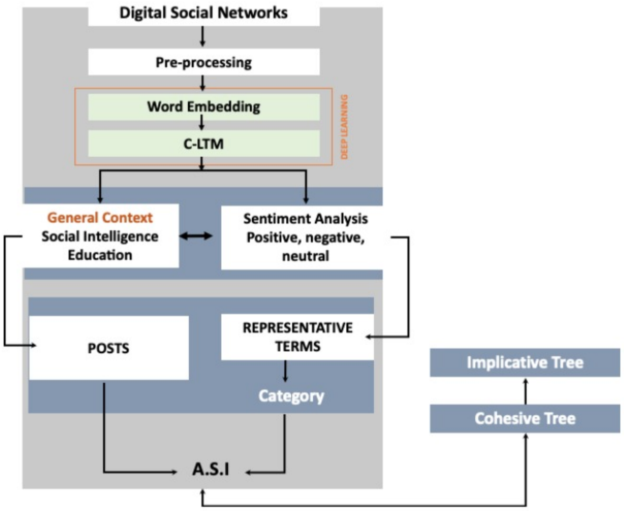

Fig. 1. Proposed Model for extracting and organizing opinion-oriented information

with the Global Vectors (GloVe) embedding technique proposed by [14] as a way of processing the data and obtaining its classification in two different aspects for the classification and identification of the type of feeling expressed in each (positive or negative). For this, in the proposed model, a C-LSTM network with a convolution of one dimension containing 100 filters of size 3 was used, followed by a layer of LSTMs composed of 300 units, afterwards 3 fully-connected layers with 300, 100 and 65 units, respectively, finally a dropout of 0.5 is added.

A C-LSTM network consists of a one-dimensional convolutional layer, responsible for extracting high-level attributes in the representation of sentences, preceding layers of LSTM that allow the obtaining of global and temporal semantics of the sentences. The Embedding layer is responsible for receiving the pre-trained weights by [52], with a corpus in Portuguese of the Global Vectors (GloVe) method used as the network’s input. With this, the network will have as input dense vectors from which it is able to process, unlike sparse data that are the texts taken from social networks.

According to the proposed approach involving DL, it was necessary to obtain a set of annotated posts in a total of 42.062 thousand publications extracted. In the training phase of the model, a dataset with 1138 publications annotated from the Web system (classification dataset) developed by [8] was used. The accuracy achieved in the test set (20% of the annotated sample) was 76.32%, regarding the classification for the education axis in the 11 axes. With the cataloging of the posts, we proceeded with step II of classifying the feeling as negative, positive or neutral, in which an accuracy of 80.70% was achieved. 9098 posts related to education were obtained.

After this stage, we started to analyze the set of co-occurring words, in order to obtain the terms representative of the contents covered for the education axis. A posteriori we organize these expressions in 32 sets of imperfect semantic synonyms, that is, when the meaning of the words is only similar and not identical. The following categories were organized: *Manager, City, Communication, Secretaries, Didactics, Training; School, University, Students, Gender, Activities, Support, Evaluation, Infrastructure, Dissemination, Educator, Work, Modalities, Curriculum, Motivation, Environment, Registration, Promotions, NGOs, Culture, Time, Citizenship, Celebrations / economy, Sport, Tradition, Popular Culture, Health and Unclassified*. After this process, the frequencies of occurrence of each term in the posts were again calculated and computed in their respective category.

For the exploration of specific contexts, it was carried out through Implicative Statistical Analysis (ASI, [53], [54], using the Hierarchical Implicative and Cohesive Classification system - CHIC (Classification Hierarchique Implicative et Cohesitive - Version 6.0, 2012). The purpose of CHIC is: *(i) to extract association rules between variables from a data set; (ii) provide a quality index for that association; and (iii) represent the structures of the variables obtained through these rules* ([55]). This mapping outlined by the researcher allows him to develop his interpretations from points of opposition or approximations and contradictions or repetitions of the data [16].

In order to represent these contexts, the software generates trees of hierarchical classifications of similarity indices and cohesion indices of implication, in addition to implicit graphs. The similarity index defines the principle of comparison between the observed and the data generated by chance, whereas a cohesion index regroups the classes of variables considering the meaning and strength of the association [55] . This force of implication of a rule evaluated by the difference between the observed number of counterexamples and the average number generated only by chance, is precisely the notion of contributions associated with a class that interests us.

For the analysis of the model, we selected the option "significant nodes", "long calculation", type of "entropic" implication, according to the "binominal law" for the confidence interval between 0.85 and 1. According to [56] , the closer to 1.0 and above 0.7 the better the data reliability. From this processing we generate hierarchical trees of similarity and cohesion.

V. Results

The analysis of the data obtained in the feeling classification phase showed that of the total of 9098 posts in Education. Of these, 86% have positive perceptions, 11% negative perceptions and 3% have neutral comments. Despite these positive education indicators, in this paper we highlight the need to understand and summarize the points listed by the population as being negative. Therefore, in this way, in [8] we seek to understand the social representations mentioned and related about the negative content of the posts.

A. Similarity Hierarchy Analysis

We interpret the similarity in order to identify in the data set proximity of answers, that is, similar opinions. For the Similarity (S) of the Education-positive axis, the confidence interval around a similarity from 1 to 0.921 was considered, distributed among the 31 levels of the tree. This level of confidence indicates a possible precision in face of a calculation made, since there is a narrow gap between the values. This proximity indicates a higher probability of the percentage of posts classified by model o for the education axis.

To better understand this index of association of similarity trees, it is possible to observe the values of occurrence, mean and standard deviation of each representative term. For example, the values displayed in table I for the terms "evaluation", "manager" allow to verify the distributive concentration of such expressions in the messages, and whose values demonstrate the regularity of their occurrence (average) and the degree of oscillation of that opinion in the data set (standard deviation).

TABLE I
Occurrence, Average and Standard Deviation Source: Authors / research

n col : 32, n fil : 6627 Variable (occurrence, average, standard deviation)
Evaluation (720.00, 0.11, 0.28); Manager (810.30 ,0.12 , 0.28); City (651.70 ,0.10 , 0.25); Secretaries(409.50 , 0.06, 0.22); School (1644.70,0.25 , 0.40); Activities (1854.00 , 0.28, 0.41); Training (351.60 ,0.05 ,0.19); Didactic (474.00 ,0.07 ,0.21); University (387.80 ,0.06 ,0.21); Motivation (543.70 , 0.08 , 0.24); Modality (180.90 , 0.03 , 0.13); Students (613.60 , 0.09 ,0.25); Infrastructure(550.50 , 0.08,0.24);Popular Culture(82.00 ,0.01 ,0.09); Curriculum(259.70 ,0.04 ,0.17);Gender (110.50 ,0.02 ,0.02, 0.11);Educator (261.50 , 0.04, 0.03); Time (293.30 , 0.04 , 0.17); Sport (715.20 ,0.11 , 0.29); Support (230.90 , 0.03, 0.15); Registration (205.10 , 0.03 , 0.15); Dissemination (239.50 , 0.04 , 0.17); Health (60.30 , 0.01,); Citizenship (219.20 ,219.20, 0.15);Culture (1439.30, 0.22 , 0.39); Promotion (163.20 ,0.02 ,0.13); Fomentation /Economy (205.30 ,0.03 , 0.15); Environment (168.60 , 0.03 , 0.13); Tradition (69.60 ,0.01 ,0.9);Unclassified (22.60 , 0.00, 0.05); NGOs(50.20 , 0.01, 0.08); Work (12.20 ,0.00 , 0.03)

When the hierarchy of categories is represented by a tree, each node at level n is associated with only one node at level n-1, so each variable is associated with another predecessor variable or with a class (set of variables). The deeper the hierarchical level of implication, the more difficult it is to predict the correct relationship between classes. This is because classes at the deepest levels represent more specific information and are produced by models induced from fewer training examples. In the tree shown in figure 2, we can see that the hierarchical classes most compatible with the values and quality of the implication are at levels 1, 4, 9, 13, 18, 21, 24, 26, 29, 31 highlighted in the image by the red lines.

Thus, we come to understand this grouping of hierarchical information through the identification of associative meta-rules (set of variables that may be related to another variable or another level of the tree). According to [16], there is consecutiveness or precedence when one variable is produced after another, but without necessarily being true that the first

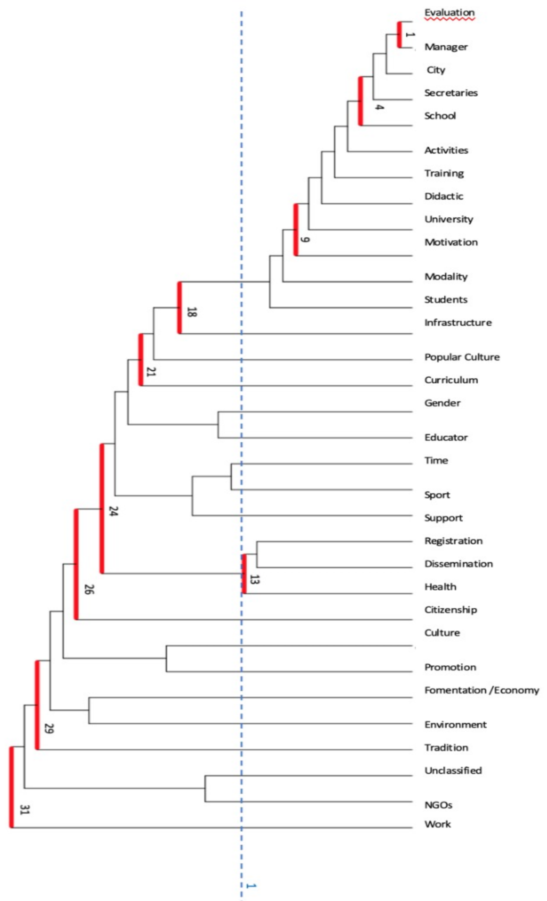

Fig. 2. Similarity Analysis - Positive Education Axis Source: Authors / research

can be produced independently of the second. Thus, these forecasts of consecutive appearances of variables and between classes of variables, allow the researcher to identify evidence of relationship. Due to the capillarity of the established relationship and the ability to predict classes, in the mapping outlined by CHIC for the Education-positive axis, we chose to distinguish globally in the hierarchy only the significant nodes (levels 1, 4, 6, 9 and 13) with values similarity implication (S = 1). The other representative terms, although related to the educational activities developed in the territory, will not be the focus of this stage of analysis. The similarity index between variables objectively shows the level of similarity between two variables or between classes of variables according to the principle of comparison between what was observed and what would be given by chance. [53].

The construction of the level 1 meta-rule (S = 1) with the representative terms (”evaluation”, ”manager”) allows us to identify an emphasis on the perception of the population regarding the role of the manager for the positive evaluation of education in the city. However, it would be worth questioning which management is referred to by the population. A probable answer to this question is made clear when we analyze the next significant node presented by the tree. For the construction of its meta-rule, level 4 (S = 1) lists the following classes of terms and representative terms (((((”evaluation”, ”manager”) ”city”) ”secretariats”) ”training”). This grouping provides evidence that the evaluated city manager is associated with the municipal education department, whose perception of capacity is also related. The meta-rule established by level 9, on the other hand, presents the relationship ((((((((((”evaluation”, ”” manager ”)” city ”)” secretariat ”)” training ”)” motivation ”)” infrastructure ”)” educator ”)” school ”)” student ”) that reveals a necessary combination between an efficient management for education that allows the student to be part of a public policy of social formation.

Level 13 lists the meta-rule ((”didactic”, ”time”) ”inscription”) in another structural branch of the similarity implication. This organization reveals that the population evaluates the impact of the teaching method, consisting of a set of processes and procedures that the teacher uses to pursue the purpose of teaching (”didactic”). As a result, the continuity of school education (”time”) represented by the student’s permanence in the school environment (”enrollments”).

It is important to highlight that the synthesis of the interaction between the elements is not the simple sum or juxtaposition of them, but the emergence of something new, previously nonexistent. The validation of data interpretation is established by a new verification step in which the analysis model is (re) applied to the set of posts classified as positive and negative, as well as the set of representative terms that made up the categories (variables) of greater similarity. The purpose of this process is to detail the quality of association of the information.

For the analysis of the Education-negative axis, the option for a verification was maintained, considering that the examination of variables with similarity index is equal to 1. Present only at levels 1, 2 and 3, highlighted in figure 3 by the red lines, and only level 1 (”manager”, ”city”) indicates level for a significant node, that is, in this association with the values are compatible with the quality of the implication. The classes gathered in the middle of levels 4 and 21 reached an interval between the values of 0.999 and 0.903, for similarity, with the others being below 0.83.

The construction of the meta-rule (figure 4) of level 1 (S = 1) with the representative terms (”manager”, ”city”) allows to identify a dissatisfaction of the population regarding the management of the city evaluated as negative for education. When comparing the results drawn by the most significant nodes in both trees (positive and negative) we found differences in perceptions regarding the evaluation of the positive for the manager and a negative evaluation of the evaluation of the city management.

This congruence of themes presented by the analysis method demonstrates an alignment between the classes gathered during the mining and data classification stages. The constituted meta-rule evidences a positive evaluation of the manager, however it is confronted by an evaluation of negative management of the city. The combination of these perceptual signs of a city

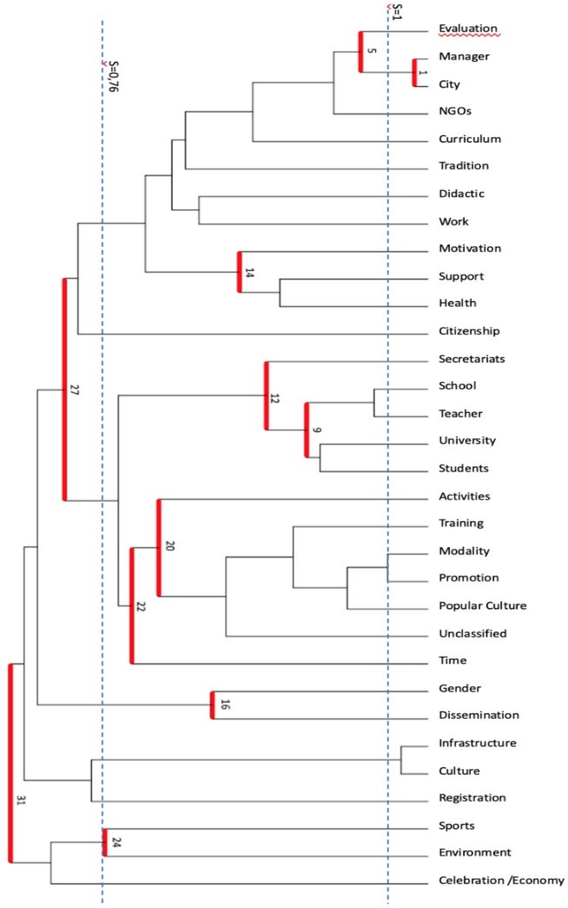

Fig. 3. Similarity Analysis - Negative Education Axis Source: Authors / research

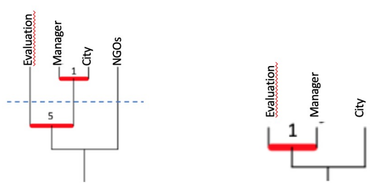

Fig. 4. Negative significant nodes / Positive significant nodes

intelligence, indicates antagonistic signs expressed by opinions. To understand this result, we launched two hypotheses based on the idea that associations refer to different levels of government: the first is the summary of negative opinions, the manifestation of an evaluation by the city manager; the second hypothesis associates positivity with the education manager in the municipality.

In both cases, there is a highlight in the opinions for the management as the main factor of perception of the population in the education axis. However, the assumption that we are facing a different management process (Mayor and Municipal Secretary of Education), needs to be better explored with a specific analysis of the posts gathered by these representative terms. Therefore, in the next session, an analysis of the speech expressed in posts associated with representative terms for the Education-negative axis will be carried out.

B. Analysis of Cohesive Hierarchy

The intra and interclass relations of the aspects indicated by the quality of the implication are presented by CHIC through the analysis of the cohesive hierarchy. According to [53], a cohesive tree graphically translates the successive fitting of the classes constituted from a decreasing value between the variables of the hierarchy. The cohesion confidence interval allows the researcher to evaluate the constitution of classes whose implicative meaning is organized according to a set of semantic associations, that is, linking the representative terms from the chain of ideas.

For the analysis of Cohesion (C) of the education-positive axis, we considered the confidence intervals equal to or greater than 0.748, distributed over the 23 levels of the cohesion tree. Since, in the first 5 classification levels, the cohesion value was equal to 1 and from level 6 to level 10, we observed a range between 0.999 and 0.907. The significant nodes, highlighted in figure **??** by the red arrows, indicate a sense of class orientation at levels 1, 6, 8, 10, 12, 14, 16. In this representation we observe that in the first level, an ordered class (”manager”¡=¿”city”) is formed whose implication of the variable (”manager”) over the variable (”city”) is the strongest among all possible implications , corresponding to him, the first significant node. Then, at level 8 of the same branch, the meta-rule emerges (”citizenship” (”manager”¡=¿”city”)), oriented as follows: if ”citizenship” is true then (if ”manager” is true then ” city ”) is usually.

This quality oriented towards the meta-rule (”manager”¡=¿”city”), indicates a mutual (equal meaning between the variables) and powerful cohesion for implication (C=1) of the association between the opinions that connect a city condition to a management condition. Seeking to complement the verification of this argument, we analyzed in the same branch the level 8 node whose correlated meta-rule (”citizenship”(”manager”¡=¿”city”)), reinforcing the role of public management for the quality of life in the city. According to [57], citizenship is the right to life in the full sense.It is a right that needs to be built collectively, not only in terms of meeting basic needs, but of access to all levels of existence. [58] points out that in the relationship between the Citizen School and the Educating City, it is necessary to establish active citizenship, this means establishing permanent channels of participation, encouraging the organization of communities so that they take control, in an organized way, in their hands city.

Resuming our problematization that asks this reference for manager’s evaluation, previously presented in the similarity for the Education-positive axis, we verified at level 6 a meaning node for the meta-rule ”training”=¿(”secretaries=¿(”educator”=¿”evaluation”)). We noticed

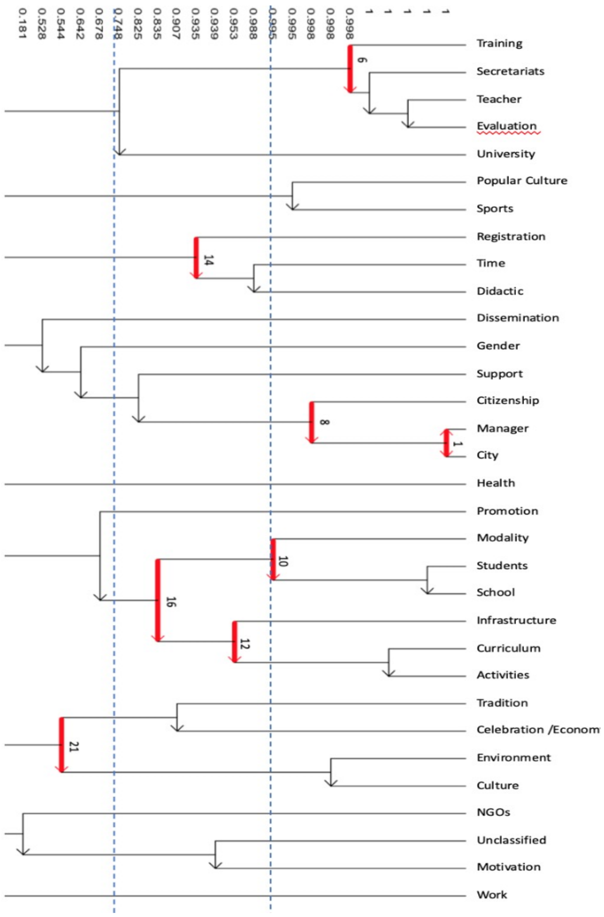

Fig. 5. Cohesive Analysis - Positive Education Axis

that the set of these variables express the management capacity linked to the work implemented by the municipal secretariats, being for this axis of analysis the Municipal Secretariat of Education. This finding reinforces our hypothesis that despite the similarity analysis for the negative feelings to relate the evaluation of the city manager and the positive feeling to an evaluation of the manager, we are not facing the same hierarchical level. The cohesion index clarifies this semantic relationship by indicating as positive the management carried out by the staff of the secretariats, as we can see in the detail in figure 6.

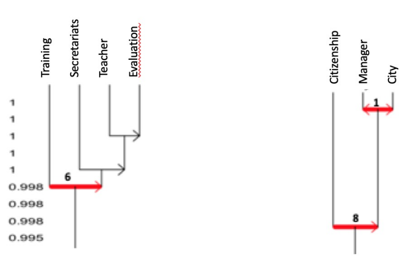

Fig. 6. Fragments - Positive Education Axis

Proceeding with the analysis of Cohesion (C) for the education-negative axis, we consider the 6 levels distributed in the confidence interval equal to or greater than 0.771, with their significant nodes distributed in levels 1, 6, 11, 16, 19 according to the emphasis of red arrows shown in figure x. As a result, the main oriented quality (C=1) reveals the meta-rule (”infrastructure”=¿”culture”), in which we perceive a negative perception of the population regarding the municipality’s infrastructure directed to cultural goods. According to [58], the structure of cities is reaching the limit of what is bearable due to its rates of violence, unemployment, lack of housing, transportation, basic sanitation, among others. The only way out, pointed out by the author, is a radical transformation of urban spaces in Educating Cities, in which innovative experiences and practices are transformed into citizen empowerment projects for the population. This means using the new training spaces created by the information society, to integrate and articulate socially significant knowledge and knowledge. There are many transformative social energies that are still dormant due to the lack of an educational perspective on the city.

This understanding of the need to revisit public infrastructure to assess its impact on the relationship between education and culture in a city is reinforced by the significant level 6 node (C=0.771). The meta-rule ((”city”=¿”manager”)”assessment”) reveals how the fragility of the public service system has an impact on the population’s perception of management assessment. When we compare the ramifications presented by both cohesive trees (positive and negative), we observe that the relationships manifested by the phenomena themselves do not present greater dependence on associated events (classes of categories), which allows us to clearly determine problematic focuses.

In the comparative detail presented by figure 8 , we verified evidence of a positive perception of the city related to the training of the secretariat in relation to the educator’s evaluation. However, when comparing the obtained values, we verified that the quality of the implication of the meta-rule (”manager”¡=¿”city”) with (C=1) cannot be compared in the same meaning for the meta-rule ((”city”=¿”manager”)) ”evaluation”) with (C= 0.771). What the analysis of the variables allows us to identify is the existence of a positive perception regarding the management of education in the city, however it is also possible to state there are problems, associated with the infrastructure capacity of the same public management as its ability to offer different educational equipment its population.

As strategies to validate the arguments established so far, implicit graphs are generated, based on the selection restricted to the representative terms that make up the analysis categories for the positive confidence meta-rule (”manager” ¡=¿ ”city”) and meta-rules negative confidence (”infrastructure” =¿ ”culture”). It is important to highlight that selecting different layers of relationship between the attributes and the analysis variables, makes it possible to identify new hierarchical flows of relationships. These results will be presented in future publications.

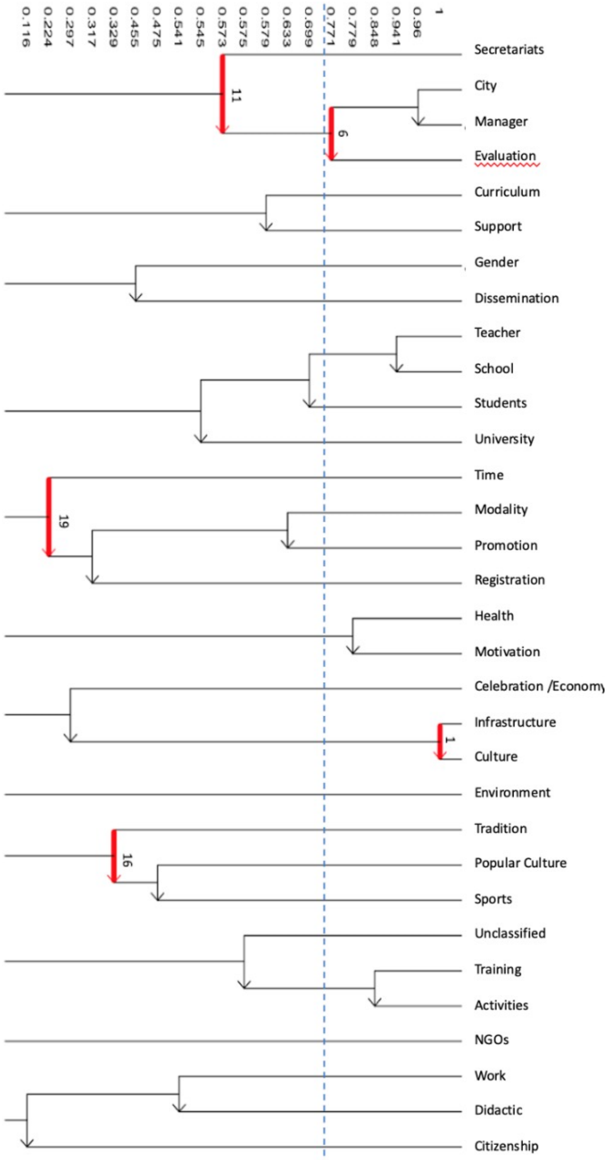

Fig. 7. Cohesive Analysis - Education-negative axis

VI. Conclusion

The study was carried out with the intention of developing a methodology that adds technical and technological conditions for visualization, organization, and construction of models in order to summarize phenomena associated with the data evidenced in this study from the concepts narrated by individuals about Education. The adaptive construction of the methods of Data Mining, Deep Learning and Implicative Statistical Analysis (ASI) represent an arrangement proposal that has been giving results in the face of the challenge of extracting these opinions and understanding their collective and spontaneous content in order to obtain indicators for decision making. decision of teachers and managers.

Besides that, the data obtained allows us to identify that the content of the posts in the midst of social networks are capable of providing evidence of representations as expressions around which society expresses its opinion. What we have tried to argue so far defends the hypothesis that there is in the mass speeches in the midst of social networks a social collectivity capable of representing a form of collective intelligence. Discovering public opinion is an important factor in the decision-making process, mainly in order to guide public authorities in their relationship with citizen participation. However, the amount of subjective content existing in textual documents on the Web makes the activity of interpreting this type of information complex, aggravated when it is necessary to extract and classify subjectivity in a large volume of data from different sources. As future work it is proposed to carry out the analysis of the implicit graph generated by the CHIC software, seeking to graphically translate different network of relationships and levels of implication between the variables.

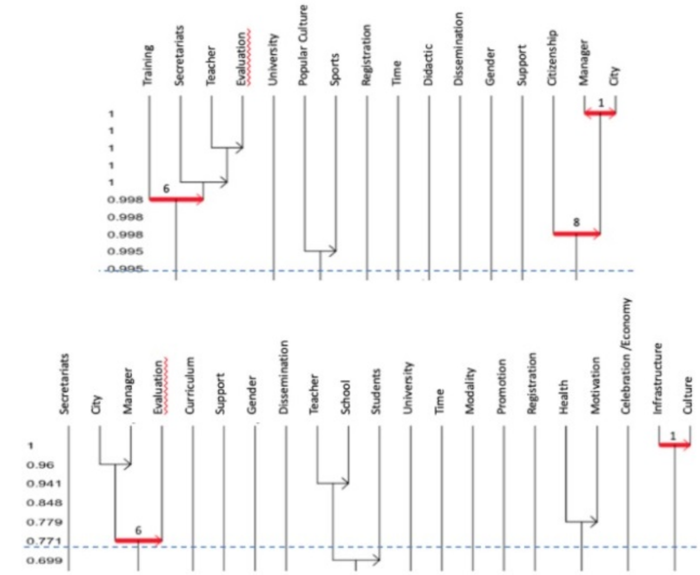

Fig. 8. Positive sense of confidence - Negative sense of confidence

Acknowledgment

Federal University of Rio Grande and Federal University of Pampa. This study was financed in part by the Personal Improvement Coordination of Higher Education (Capes) Finance Code 001.

References

[1] C. R. Miller, "Public opinion polls and public schools," *Teachers College Record*, vol. 43, pp. 245–54, 1942.

[2] W. D. Blum, "Opinion toward education in montreal, canada," *The Journal of Experimental Education*, vol. 15, no. 4, pp. 219–267, 1947.

[3] A. M. Rose, Lowell C e Gallup, "A 32ª pesquisa anual phi delta kappa / gallup sobre as atitudes do público em relação às escolas públicas," vol. 82, pp. 41–58.

[4] W. H. Woodall, M. J. Zhao, K. Paynabar, R. Sparks, and J. D. Wilson, "An overview and perspective on social network monitoring," *IISE Transactions*, vol. 49, no. 3, pp. 354–365, 2017.

[5] T. H. McCormick, H. Lee, N. Cesare, A. Shojaie, and E. S. Spiro, "Using twitter for demographic and social science research: tools for data collection and processing," *Sociological methods & research*, vol. 46, no. 3, pp. 390–421, 2017.

[6] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*, pp. 415–463, Springer, 2012.

[7] V. Patel, G. Prabhu, and K. Bhowmick, "A survey of opinion mining and sentiment analysis," *International Journal of Computer Applications*, vol. 131, no. 1, pp. 24–27, 2015.

[8] V. Kwecko, F. Tolêdo, F. Mota, G. Almeida, S. Devincenzi, F. Z. Ferreira, A. Bicho, and S. S. da Costa Botelho, "Ciência de dados aplicada na análise de processos cognitivos em grupos sociais: um estudo de caso," in *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, vol. 29, p. 1543, 2018.

[9] C. C. Aggarwal and C. Zhai, *Mining text data*. Springer Science & Business Media, 2012.

[10] A. Valdivia, M. V. Luzón, and F. Herrera, "Sentiment analysis in tripadvisor," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 72–77, 2017.

[11] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.

[12] E. W. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision support systems*, vol. 50, no. 3, pp. 559–569, 2011.

[13] A. Moreno, P. Rita, and J. Guerreiro, "Sentiment analysis in online reviews classification using text mining technique," *Sentiment analysis in online reviews classification using text mining technique*, 2019.

[14] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[15] S. Abujar, M. Hasan, and S. A. Hossain, "Sentence similarity estimation for text summarization using deep learning," in *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, pp. 155–164, Springer, 2019.

[16] S. A. Almouloud, "O que está pode detrás do chic," *Uso do CHIC na Formação de Educadores: à guisa de apresentação dos fundamentos e das pesquisas e foco. José Armando Valente*, 2015.

[17] V. M. Netto, *Cidade & Sociedade: as tramas da prática e seus espaços*. Editora Sulina, 2014.

[18] C. S. d. Freitas, F. F. Lima, and F. Q. Lima, "Os desafios ao desenvolvimento de um ambiente para participação política digital: o caso de uma comunidade virtual legislativa do projeto e-democracia no brasil," *Organizações & Sociedade*, vol. 22, no. 75, pp. 639–658, 2015.

[19] C. S. Steinberg, "Mass media and communication," 1966.

[20] G. Tarde and J. Veloso, *A opinião ea multidão*. 1991.

[21] N. M. de Sena, "Espaço público, opinião e democracia," *Que Sais-Je*, p. 3, 1977.

[22] M. Augras, *Opinião pública: teoria e prática*. Vozes, 1974.

[23] S. Lundåsen, "Podemos confiar nas medidas de confiança?," *Opinião pública*, vol. 8, no. 2, pp. 304–327, 2002.

[24] P. Magalhães, *Sondagens, eleicões e opinião pública*. Fundacão Francisco Manuel Dos Santos, 2016.

[25] J. C. S. Sampaio, "Proposição de um modelo de retroalimentação da gestão do processo de projeto a partir de medições de satisfação de clientes," 2010.

[26] R. Figueiredo and C. Coutinho, "A eleição de 2002," *Opinião Pública*, vol. 9, no. 2, pp. 93–117, 2003.

[27] P. H. Smith and M. R. Ziegler, "Democracias liberal e iliberal na américa latina," *Opinião Pública*, vol. 15, no. 2, pp. 356–385, 2009.

[28] G. I. Nitin, G. Swapna, and V. Shankararaman, "Analyzing educational comments for topics and sentiments: A text analytics approach," in *2015 IEEE frontiers in education conference (FIE)*, pp. 1–9, IEEE, 2015.

[29] S. E. Sorour, T. Mine, K. Goda, and S. Hirokawa, "Predicting students' grades based on free style comments data by artificial neural network," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, pp. 1–9, IEEE, 2014.

[30] C. Leston-Bandeira and R. Cabral, "O parlamento ea opinião pública em portugal: argumentos para reforma," *Análise Social*, pp. 467–482, 2003.

[31] H.-C. Lin, P. F. Bruning, and H. Swarna, "Using online opinion leaders to promote the hedonic and utilitarian value of products and services," *Business Horizons*, vol. 61, no. 3, pp. 431–442, 2018.

[32] A. P. Pross, *Group politics and public policy*. Oxford University Press, 1992.

[33] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, 2017.

[34] S. Hussain *et al.*, "Survey on current trends and techniques of data mining research," *London Journal of Research in Computer Science and Technology*, 2017.

[35] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, p. 1, 2015.

[36] S. Golder and M. Macy, "Social science with social media," *ASA footnotes*, vol. 40, no. 1, p. 7, 2012.

[37] C. Brickman Bhutta, "Not by the book: Facebook as a sampling frame," *Sociological methods & research*, vol. 41, no. 1, pp. 57–88, 2012.

[38] N. Heaivilin, B. Gerbert, J. Page, and J. Gibbs, "Public health surveillance of dental pain via twitter," *Journal of dental research*, vol. 90, no. 9, pp. 1047–1051, 2011.

[39] J. e. L. C.-H. Naamã, Mor e Boase, "É realmente sobre mim? conteúdo da mensagem nos fluxos de conscientização social," in *Anais da conferência de 2010 da ACM sobre trabalho cooperativo apoiado por computador*, pp. 189–192.

[40] U.-D. Reips and P. Garaizar, "Mining twitter: A source for psychological wisdom of the crowds," *Behavior research methods*, vol. 43, no. 3, p. 635, 2011.

[41] E. Hwang, *Suicídio por contágio e a comunicação midiática*. PhD thesis, Universidade de São Paulo, 2018.

[42] D. M. Ferreira and L. Mattos, "O efeito contágio da crise do subprime no mercado acionário brasileiro," *Encontro Nacional de Economia*, 2012.

[43] B. Bertolino, M. Kiener, and R. Patterson, *The Therapist's Notebook on Strengths and Solution-Based Therapies: Homework, Handouts, and Activities*. Taylor & Francis, 2010.

[44] A. L. Holbrook and J. A. Krosnick, "Social desirability bias in voter turnout reports: Tests using the item count technique," *Public Opinion Quarterly*, vol. 74, no. 1, pp. 37–67, 2010.

[45] C. A. Cassa, R. Chunara, K. Mandl, and J. S. Brownstein, "Twitter as a sentinel in emergency situations: lessons from the boston marathon explosions," *PLoS currents*, vol. 5, 2013.

[46] C. Sutton, A. McCallum, *et al.*, "An introduction to conditional random fields," *Foundations and Trends® in Machine Learning*, vol. 4, no. 4, pp. 267–373, 2012.

[47] J. A. Koepfler and K. R. Fleischmann, "Studying the values of hard-to-reach populations: content analysis of tweets by the 21st century homeless," in *Proceedings of the 2012 iConference*, pp. 48–55, 2012.

[48] N. A. Diakopoulos and D. A. Shamma, "Characterizing debate performance via aggregated twitter sentiment," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1195–1198, 2010.

[49] M. Castells, *Another economy is possible: Culture and economy in a time of crisis*. John Wiley & Sons, 2017.

[50] G. King, R. O. Keohane, and S. Verba, *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press, 1994.

[51] C. Zhou, C. Sun, Z. Liu, and F. Lau, "A c-lstm neural network for text classification," *arXiv preprint arXiv:1511.08630*, 2015.

[52] N. Hartmann, E. Fonseca, C. Shulby, M. Treviso, J. Rodrigues, and S. Aluisio, "Portuguese word embeddings: Evaluating on word analogies and natural language tasks," *arXiv preprint arXiv:1708.06025*, 2017.

[53] R. Couturier, A. Bodin, and R. GRAS, "A classificação hierárquica implicativa e coesiva," *Manual Curso CHIC versão*, vol. 2, 2004.

[54] R. Couturier, A. Bodin, and R. Gras, "Chic version 3.5," *Orléans: ARDM*, 2005.

[55] R. Gras and S. A. Almouloud, "A implicação estatística usada como ferramenta em um exemplo de análise de dados multidimensionais," *Educ Mat Pesqui*, vol. 4, no. 2, pp. 75–88, 2002.

[56] G. Martins and C. R. Theóphilo, "Metodologia da investigação cientifica," *São Paulo: Atlas*, pp. 143–164, 2009.

[57] J. F. Covre *et al.*, "Contribuições da musicoterapia para a comunicação de crianças com alterações da linguagem," 2015.

[58] M. Gadotti, "A escola na cidade que educa," *Cadernos Cenpec— Nova série*, vol. 1, no. 1, 2006.