Comparative analysis of the government plans of the Peruvian presidential candidates, SDO(UN) and State Policies of the National Agreement based on NLP

Honorio Apaza Alanoca¹, Josimar Chire² and Jimy Oblitas³

¹ Data Science Research Group
² Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP), São Carlos, SP, Brazil
³ Facultad de Ingeniería, Universidad Privada del Norte, Cajamarca, Perú
hapazaas@unam.edu.pe, jecs89@usp.br, jimy.oblitas@upn.edu.pe

Abstract. The analysis of government proposal during elections from political parties is vital to choose the next authorities in any city or country. In this paper, we use a text mining approach to analyze the documents and provide an easy visualization to support an easy analysis. Besides, a comparison with a national plan based on sustainable development objectives of UN(United Nations) from 2030 Agenda is performed using Natural Language techniques.

Keywords: Natural Language Processing, Text Mining, Data Science, System Recommender, Elections, Politics, Peru, South America

1 Introduction

Election of authorities is an important event, because citizens will choose the people who will represent them and purpose projects to improve the national, regional context. Traditionally, political parties promote their candidates through mass media, i.e. radio, television, social networks and more. Candidates travel to visit cities and gain more electors.

In Peru, to participate in president elections is a requirement to send a government proposal or plan to Jurado Nacional de Elecciones (National Elections Jury). This document summarizes the proposal of the candidates, considering the most important problems for the party and solutions that they purpose. Usually, these documents have dozens of pages and these are not read for citizens to choose the next authority. Besides, United Nations (UN) purposed an 2030 Agenda to summarize the most important issues which need special attention for governments related to poverty, communication, discrimination and more.

In 2015, the United Nations (UN) adopted a new international development agenda: the 2030 Agenda that includes the 17 Sustainable Development
Goals and 169 targets. This agenda specifies the need for actions to strengthen sustainable economic growth, decent employment and industrialization in all countries [Caribbean, 2017].

The 2030 Agenda considers a complex combination of fairly detailed thematic targets, through a comprehensive approach that requires addressing sustainable development as a necessary integration of the social, economic and environmental axes [Nieto, 2017]. Although it is recognized that each country has its own priorities, this agenda is a reference for government plans seeking an adequate sustainable development of Peru. Therefore, measuring the alignment or possible evolution of government plans of presidential candidates is a necessary task.

In this context, the use of software tools, such as text mining, emerges as a quick and interesting proposal to measure trends. In addition to the fact that, in the Peruvian context, such tools are not used yet, this contrasts with global trends in the use of software tools that are already established, as in the campaigns of Trump and Bolsonaro, in the United States (USA) and Brazil, which illustrate policy facts that have been favored by ICTs [Garcia-Nunes et al., 2020].

Natural language processing has shown potential as a promising tool to exploit urban data sources. Authors, such as [Cai, 2021], suggest that the use of urban big data sources is still starting and the most studied areas are: urban governance and management, public health, land use and functional zones, mobility and urban design, having been very useful in expanding study scales and reducing research costs.

Text Mining area uses a well-know Data Mining approach, from Data Collection, Exploration, Analysis to Visualization. Text Mining focuses in Text Analysis, uses Natural Language Techniques (NLP). Many studies were performed to analyze different problems from different areas, i.e. epidemiology [Chire Saire and Oblitas Cruz, 2020], politics [Sharma and Shekhar, 2020], marketing, etc.

Applications of Text Mining in Politics and Elections, i.e. Anticipating Political Behaviour [Sangar et al., 2013], Study Voting Patterns [Bagui et al., 2007], Fraud Identification [Poloni and Formolo, 2015], Sentimental Analysis of citizens [Sharma and Ghose, 2020], Election Result Prediction [Ramteke et al., 2016] and more.

The objective of this paper is analyze the government proposal of Peruvian candidates to president elections using a Text Mining Approach to support an easy understanding of the documents. Besides, perform a matching process with national plan adapted from 2030 Agenda, to check how important are these objective for political parties.

Section I includes the review of the bibliography, Section II develops the work proposal, Section III discloses the results of the research and in Section IV gives conclusions, last section presents future work.
2 Proposal

Natural language processing is a process transformation the text information in numeric data [Di Giuda et al., 2020]. This work is based on the following research process:

![Flowchart of research process](image)

Fig. 1: Research process, this process is planed and used for [Kim et al., 2017]

2.1 Data collection

For the present work, 18 government plans of the candidates for the presidency of the Republic of Peru have been collected. Also the sustainable development goals and policies of the state of the national agreement, the sustainable development goals (SDGs) promoted by the United Nations, whose predecessor are the Millennium Development Goals, constitute an inclusive global agenda with goals for 2030 [secretaria ejecutivo del acuerdo nacional, 2017].

2.2 Data analysis

Jaro Winkler is the main algorithm to perform comparative text analysis of documents (government plans of the candidates) with the Sustainable Development Goals (SDGs) promoted by the United Nations.

\[
Sim_j(s_1, s_2) = \begin{cases} 
0 & \text{if } m = 0 \\
\frac{1}{3} \left( \frac{m}{s_1} + \frac{m}{s_2} + \frac{m-t}{m} \right) & \text{otherwise}
\end{cases}
\]

The objective is to calculate the distance of the strings of texts that are written in the plans of the government of the candidates and the objectives and policies of sustainable development of the state of the national agreement. In this first preliminary test of the research we are interested in knowing what results are obtained with Jaro Winkler.

2.3 Reporting

Finally, the last stage of the research is to make a report on the results obtained, in this case the results are the Jaro Winkler distance between the plans of the candidates’ government and the objectives and sustainable development policies of the state of the national agreement.
3 Results

This section shows the result frequency of terms in a word cloud, it can be seen that each candidate highlights a particular topic, such as: System, Health, Program, etc. This result is due to the fact that currently the nation and the world are suffering from a global pandemic, therefore, the plans of the candidates’ government propose proposals to solve problems related to health. This also shows that other important issues such as education, economics, etc. have been neglected. Especially issues related to sustainable development goals (SDG) promoted by the United Nations.

Among the candidates’ plans, the one that stands out the most is the government plan of the political party Avanza Pais on the economic issue. It can also be seen that the Accion Popular political party has a uniform distribution in its government plan on the issues of economy, health, education and politics. Can be seen in Figure 3.

In this case we can vary the issues we want to measure, this can be according to the context of the moment and different sectors of society, they have different problems and needs, so it is important to analyze from other points of view, social classes and thoughts.

Below we present a graphical (Figure 4) representation of how similar are the government plans of the candidates for the presidency of Peru, in Figure 4 it can be seen that they are not so identical, but if you can see the degree of similarity they have, but This is due to the fact that government plans clearly address very similar issues that translate into social problems (health, economy, programs, etc.) and government (judiciary, corruption, congress, etc.).

In the experiment, the differences by prolific class were also denoted, in some cases the distance is very noticeable between the political parties considered to
be on the left with those on the right. Which can be similar in the daily exercise, which obviously have very different thoughts, therefore very different proposals between these two sides of Peruvian politics.
In this section we are going to analyze the distance between chains of texts written in the government plans of the candidates and the objectives and policies of sustainable development of the state of the national agreement, we try to differentiate the similarities between these two documents, when a chain of texts is similar to another means that the document contains texts similar to the other. So, we could say that a government plan addresses one or many sustainable development goals and policies of the state of the national agreement.

In the graph above, it can be seen that the government plan of the political party Avanza Pais addresses much more than others the goal of peace, justice and solid institutions (paz justicia e instituciones solidas), followed by the political party Renovacion Ppular. However, little is addressed the objectives such as: underwater life(Vida submarina), health and well-being(salud y bienestar), end of poverty(fin de la probreza), etc.

4 Conclusions

The algorithm Jaro Winkler based on measuring the distance of text chains shows us that we are very interesting preliminary results, it shows us some differences between the government plans of the candidates for the presidency of Peru, as well as the objectives of the Sustainable Development Goals and the State Policies of the National Agreement. However, these results can be further refined with the most advanced artificial intelligence methods or algorithms.
In the present we want to highlight the way in which the differences between the government plan documents can be graphically demonstrated, this way of showing the document differences is very important for the electorate, because without having to read all the government plans, they can obtain a more general vision graphically.

5 Future work

One of the future jobs is to experiment with highly advanced artificial intelligence techniques in the discipline of natural language processing and text mining. It would be very interesting to study and experience how coherent the arguments of the candidates are in the debate with their government plan. Because there must be coherence of ideas between the proposals that are written in the government plan with what the candidate expresses in the debate, interviews in the press, etc.

References

Bagui et al., 2007. Bagui, S., Mink, D., and Cash, P. (2007). Data mining techniques to study voting patterns in the US. Data Science Journal, 6(0):46–63.
Cai, 2021. Cai, M. (2021). Natural language processing for urban research: A systematic review. Helinon, 7(3):e06322.
Caribbean, 2017. Caribbean, E. C. f. L. A. a. t. (2017). 2030 agenda for sustainable development. Last Modified: 2017-06-28T13:23-04:00 Publisher: CEPAL.
Chire Saire and Oblitas Cruz, 2020. Chire Saire, J. and Oblitas Cruz, J. (2020). Study of Coronavirus Impact on Parisian Population from April to June using Twitter and Text Mining Approach. pages 242–246.
Di Giuda et al., 2020. Di Giuda, G. M., Locatelli, M., Schievano, M., Pellegrini, L., Pattini, G., Giana, P. E., and Seghezzi, E. (2020). Natural Language Processing for Information and Project Management, pages 95–102. Springer International Publishing, Cham.
Garcia-Nunes et al., 2020. Garcia-Nunes, P. I., Rodrigues, P. A., Oliveira, K. G., and da Silva, A. E. A. (2020). A computational tool for weak signals classification – Detecting threats and opportunities on politics in the cases of the United States and Brazilian presidential elections. Futures, 123:102607.
Kim et al., 2017. Kim, K., joung Park, O., Yun, S., and Yun, H. (2017). What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. Technological Forecasting and Social Change, 123:362–369.
Nieto, 2017. Nieto, A. T. (2017). CRECIMIENTO ECONÓMICO E INDUSTRIALIZACIÓN EN LA AGENDA 2030: PERSPECTIVAS PARA MÉXICO. Problemas del Desarrollo, 48(188):83–111.
Poloni and Formolo, 2015. Poloni, Y. T. and Formolo, D. (2015). Data mining to identify fraud suspected on electronic elections. In 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems, pages 19–23.
Ramteke et al., 2016. Ramteke, J., Shah, S., Godhia, D., and Shaikh, A. (2016). Election result prediction using twitter sentiment analysis. In 2016 International Conference on Inventive Computation Technologies (ICICT), volume 1, pages 1–5.
Sangar et al., 2013. Sangar, A. B., Khaze, S. R., and Ebrahimi, L. (2013). Participation anticipating in elections using data mining methods.
secretaria ejecutivo del acuerdo nacional, 2017. secretaria ejecutivo del acuerdo nacional (2017). Objetivos de desarrollo sostenible y políticas del estado del acuerdo nacional.
Sharma and Ghose, 2020. Sharma, A. and Ghose, U. (2020). Sentimental analysis of twitter data with respect to general elections in india. *Procedia Computer Science*, 173:325–334. International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020.
Sharma and Shekhar, 2020. Sharma, A. and Shekhar, H. (2020). Intelligent Learning based Opinion Mining Model for Governmental Decision Making. *Procedia Computer Science*, 173:216–224.