Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0016.v1

Article

Unstructured Data Analysis for Risk Management of Electric Power Transmission Lines

Lucas H. Pereira[1], Rafael B. Pereira[2], Pedro H. S. Prado[3], Felipe D. Cunha[4], Fabrício Góes[5], Roger S. Fiusa[6] and Lorrany Fernanda Lopes da Silva[7]

1 Pontifical Catholic University of Minas Gerais, Department of Electrical Engineering, Minas Gerais, Brazil; lucashpereira.s@gmail.com
2 Federal University of Minas Gerais, Department of Electrical Engineering, Minas Gerais, Minas Gerais, Brazil; bambirrarafaelp@gmail.com
3 AVSystemGeo, Minas Gerais, Brazil; pedro@avsystemgeo.com.br
4 Pontifical Catholic University of Minas Gerais, Department of Computer Science, Minas Gerais, Minas Gerais, Brazil; felipe@pucminas.br
5 University of Leicester, Department of Computing and Mathematical Sciences, UK;fabricio.goes@leicester.ac.uk
6 Equatorial Energia, Brasília, Brazil; roger.fiusa@equatorial-t.com.br
7 Equatorial Energia, Brasília, Brazil; lorrany.silva@equatorial-t.com.br
* Correspondence: Fabrício Góes (e-mail: fabricio.goes@leicester.ac.uk).
† This work was supported by the project R&D ANEEL - PD-05456-0001/2019, AVSystemGeo, Equatorial, CNPq, FAPEMIG, CAPES, UFMG and PUC Minas.

Abstract: Risk management of electric power transmission lines requires knowledge from different areas such as environment, land, investors, regulations, and engineering. Despite the widespread availability of databases for most of those areas, integrating them into a single database or model is a challenging problem. Instead, in this paper, we use a single source, the Brazilian National Electric Energy Agency's (ANEEL) weekly reports, which contains decisions about the electrical grid, comprising most of the areas. Since the data is unstructured (text), we employed NLP techniques such as stemming and tokenization to identify keywords related to common causes of risks provided by an expert group on energy transmission. Then, we used models to estimate the probability of each risk. Our results show that we were able to estimate the probability of 97 risks out of 233.

Keywords: natural language processing, risk management, transmission lines, unstructured data.

---

1. Introduction

Risk management is a set of structured procedures, which aims at handling undesired outcomes, opportunities in a designed and systemic way. In the electric energy sector, risk management is present in energy trading [1,2], operation and dispatch of power plants [3–5] and energy asset management [6–8] (e.g. transmission lines).

Implementing electric power transmission lines comprises different phases, such as pre-auction, feasibility, implementation, and operation. Each phase can be affected by multiple factors related to the environment, land, investors, regulations, and engineering. In a project, each factor can be associated with different risks, and those risks probabilities can vary depending on the current phase [9,10].

In the last decade, a large amount of information was made available on the web to help managers make decisions based on data analysis. These databases come from various sources such as weather forecasts, energy regulatory reports, social media, energy market shares, concession contracts, and internal management systems companies. It is a daunting task to combine data from all those different sources into a unique model to estimate the probability of potential risks to the implementation of transmission lines since data usually come in different formats.

Data can be structured, for example, in dashboards monitoring transmission projects, where they present information to monitor compliance with concession contracts. They can also be presented in an unstructured way, as texts with information related to decisions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0016.v1

about transmission projects. An essential unstructured database is the weekly reports made available by the Brazilian National Electric Energy Agency (ANEEL) on its website [1]. Although it is challenging to extract accurate information from this database due to its unstructured nature, it is a single source that comprises most phases and areas involved in the management of transmission lines. Thus, it relieves the burden of integrating multiple database sources into a single risk management model.

This paper uses NLP (Natural Language Processing) to extract keywords from ANEEL's weekly reports and models to estimate the probability of risks in implementing electric power transmission lines based on the National Electric Energy Agency (ANEEL - Brazilian regulator) weekly reports. To the best of our knowledge, this is the first work to use weekly reports to estimate the probability of risks in the implementation of transmission lines. After extracting the data from ANEEL's website, we employed NLP techniques such as stemming and tokenization to identify keywords related to common causes of risks. A company expert on energy transmission provided the list of potential risks and causes in different phases and areas. Our results show that we could predict the probability of 97 risks out of 233 listed by the company using our NLP approach.

This article is organized as follows: Section 2 presents the background for understanding the electric power transmission line implementation. Section 3 discusses the related work. Then, Section 4 introduces the research methodology. Section 5 presents and discusses the experimental results. Finally, Section 6 presents the conclusion and new directions for this research.

2. Background

This section will describe the importance of the energy sector in social development, the stages of the life cycle of transmission projects, and which interest areas are involved in each phase of this cycle. Finally, we will explain the importance of risk management during the implementation project.

The energy sector has a crucial role in society and the countries' social and economic development. Around 70% of the energy generation in Brazil is based on hydroelectric plants, followed by thermo-electric (28%) and wind (1%) power plants. Regardless of the source, electricity flows through the transmission network with a nominal voltage above 230 kV. When it reaches substations located in cities, the voltage is lowered, and then, through a system consisting of wires, poles, and transformers, it is sent to the consumer at 127 or 220 volts.

The national generation and transmission network is interconnected and managed by the National Electric System Operator (ONS). The entity's fundamental role is to control water supplies at hydroelectric plants and forecast river flows. The National Electric Energy Agency (ANEEL) regulates this sector.

2.1. Phases of the Energy Transmission Segment

The life cycle of transmission projects goes through concession auctions carried out by ANEEL and later through the stages of the environmental licensing process, which will aim to ensure environmental feasibility and mitigation of impacts arising from the installation of these projects. Once these steps have been addressed, the projects can be implemented and executed.

The stages and requirements of environmental licensing depend on several factors, such as size, location, impacts, and environmental agencies' authorization. The following environmental licenses are required:

Prior License (PL): it is granted in the preliminary planning phase of the project. It assesses the location, environmental feasibility and establishes the essential requirements and conditions to be met in the subsequent phases of the project implementation.

[1] https://antigo.aneel.gov.br/pautas-e-atas

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0016.v1

Installation License (IL): it authorizes the installation of the project following the specifications contained in the approved plans, programs, and projects, including environmental control measures and other conditions.
Operating License (OL): it authorizes the Operation of a project, after verifying the effective fulfillment of what is stated in the previous licenses, with the environmental control measures and conditions determined for the Operation.

A project to implement transmission lines follows four distinct phases according to the environmental phases: Pre-auction (before PL), Viability Study (after PL), Implementation (after IL), and Operation (after OL).

2.2. Areas of Interest for Risk Management

Transmission projects are complex and involve integrating different areas, not only environmental issues. In this research, we are interested in the following areas:
Environment: it comprises all aspects of environmental licensing itself, such as the various impacts on physical, biological, and socioeconomic environments.
Land: it is responsible for all registration, evaluation, negotiation, indemnification, law, and regularization of properties that are necessary or, in some way, impacted by the project.
Engineering: it comprises topography and drilling studies, through the elaboration of the Basic Engineering Project, supply and electromechanical assembly, civil construction, and commissioning.
O&M Engineering (Operation and Maintenance): it involves all actions to monitor the Operation and carry out the necessary maintenance to ensure the availability of the systems and the lowest possible discount on the variable portions.
Regulatory: it comprises the obligations related to ANEEL Concession Contracts, ONS Network Procedure, in addition to technical and works safety standards.
Investors: it comprises the preparation of the business plan, the control of project costs, and decision-making based on these costs, in addition to the company's image and obligations with the stock market.

2.3. Risk Management

Identifying potential risks and their consequences is an essential project task, as it allows for the exploration of the space of uncertainty and allows better decision-making. For the identification, measurement, and assessment of the risks associated with any project, some tools can be used [11–14]. There are risks associated with all phases and areas in a transmission line implementation project. For each risk assessed in the construction and operation of transmission lines, the probability and severity of the event are the main factors to measure the line's risk level. Many variables from different sources influence their probabilities. The weekly reports from ANEEL provide insightful information that can influence many of those risks since they report various decisions taken about the transmission network.

We have identified 233 risks, with the help of an electric energy company's specialist, that are related to transmission lines implementation projects. To mention a few, we present a shortlist of the risks:

i) Non-approval of the study of the regions with indigenous settlements
ii) Increase in the number of malaria cases registered
iii) Political changes in local governments
iv) Entrepreneur's financial health weakened
v) Variation in the basic interest rate provided in the business plan for financing the project
vi) Embargoes for environmental reasons for long periods
vii) Existence of unresolved environmental issues

3. Related Work

This section presents an overview of the related work on risk management in the implementation of energy power transmission lines.

In work presented in [15], the authors point out that uncertainty can impose substantial financial risks on energy distributors. Therefore, they propose to evaluate the financial risk and the main parameters that affect the energy supply. For this purpose, a model was developed to optimize post-failure actions, minimizing the interruption costs for customers. The authors used the parameters of financial risk indices, volatility index, value at risk, and conditional risk value to quantify the risk. Finally, the proposed method is applied to a test system, with different sensitivity analyses being carried out to identify the main parameters that affect the distribution.

Aiming to improve the accuracy of failure analysis in power distribution networks, in [16] the researchers apply data mining, data cleaning, data transformation, and data integration techniques. Then, they performed a related factor analysis to warn the distribution network's risks.

In the same direction, in [17], the authors use data science techniques, such as data mining, *machine learning* and data visualization. They develop a *big data platform* that makes it possible to integrate and combine data from various sources to solve problems in the production of distribution networks. The result shows that the reliability of the power supply can be improved and the operational costs of the distribution networks significantly reduced by using data mining techniques.

Also, in [18], the authors present a model of operation management and maintenance of the power distribution network. The authors extract a one-dimensional failure features array using the K-means clustering algorithm. They use the Apriori algorithm to mine association rules for different failure modes and establish a fundamental performance matrix. Finally, the characteristics of uni-dimensional and multidimensional failures are combined based on the theory of evidence to obtain the diagnostic criteria of failure.

Considering the use of renewable natural resources in electric energy generation, the work proposed in [19] presents the need to expand the energy transmission network in Brazil, either by interconnecting areas not yet connected to the system or by raising the reliability of existing ones. This work exploits renewable energy to optimize costs and make projects more efficient.

In the work presented in [20], the authors developed a Business Intelligence (BI) environment to analyze critical data to obtain new insights about businesses and markets, aiming at improving products and services by achieving better operational efficiency and promoting the relationship with customers. For this purpose, research activities were characterized in three segments: (a) big data analysis, (b) text analysis, and (c) network analysis. The paper purposes a review of the latest generation techniques and models, summarize businesses' analyses and determine essential issues to be addressed in future research.

Also, in [21] the authors present tasks such as acquisition, mining, data refinement, and pattern recognition in decision making. The proposed architecture works on the uncertainties of climatic conditions, energy consumption, and the stock market's risk analysis. A stochastic scheduling approach that includes the mean and variance of the energy cost is considered in the optimization process to deal with these uncertainties. Data mining algorithms are applied to reduce the vast amount of raw data and recognize patterns for analysis. Lastly, the work results illustrate the efficiency of the proposed system for different case studies.

In [22], a big data analysis approach enables refinery managers to show new cause-effect correlations in adverse environmental events, immediate and root causes, areas of refinery involved in the adverse event, risk index, and corrective actions.

As in [21], we also developed a data visualization system that enables the analysis and risk management of transmission lines. The proposed system collects, analyses, and organizes data using data mining techniques similarly to [18] to reduce financial risks in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0016.v1

the implementation of transmission lines as in [15]. To the best of our knowledge, using NLP techniques to extract data from weekly reports of an electricity power regulator and estimate the risk probabilities is a novel approach.

4. Methodology

This research is divided into four steps shown in Figure 1: i) web crawling; ii) risk and causes identification; iii) risk keywords mapping; iv) risk probability estimation.

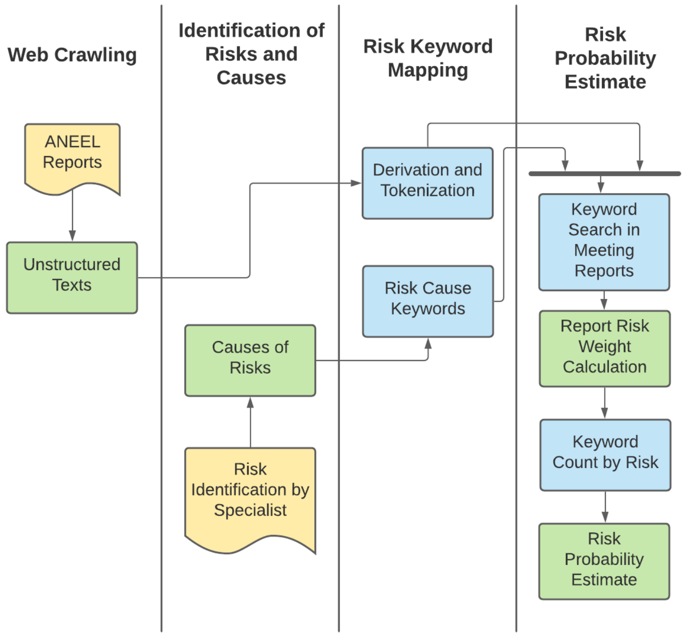

Figure 1. Diagram representing the methodology steps.

4.1. Web Crawling

ANEEL provides weekly reports available publicly on its website. These reports provide information on decisions made by ANEEL regarding the management of transmission lines. They are unstructured text files organized into paragraphs and publication dates. There is no API to extract the data, and we used web crawling to download all the reports' content from January 2005 to June 2021.

We developed a script (PHP web-based language) to access a list of pages with the weekly reports on ANEEL's website. Each list item is composed of the publication date, title, and a link to the actual report. The script downloads the corresponding HTML page for each report and extracts the report content.

4.2. Risk and Causes Identification

Based on specialist knowledge, 233 risks were identified during the acquisition and implementation of energy transmission lines. These risks are related to weather and environmental conditions, land constraints, costs, investments, etc. A specialist wrote down a list of possible causes related to the specific risk for each risk. For instance, the risk of non-approval of the Final Report of the Basic Indigenous Environmental Plan (PBAI) by FUNAI[2], related to the environmental area of interest during the Implementation phase, and its leading cause is the execution of activities together of Indigenous Communities (CIs) in disagreement with the guidelines defined in the PBAI.

[2] Justice's National Foundation for Indigenous Affairs (FUNAI).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0016.v1

Considering the risk and causes identified by the specialist, we created a dictionary with the keywords extracted from these causes. Those keywords were stemmed and tokenized along with the reports from ANEEL.

4.3. Risk Keywords Mapping

To map risks and reports, we searched the risk causes keywords in all reports and created a map linking the reports to one or more risks if at least a keyword related to that risk was found on a report. A risk could be linked to zero or more reports by the end of this step.

4.4. Risk Probability Estimation

To estimate the risk probability, we use an approach based on the TF-IDF statistical technique (Term Frequency-Inverse Document Frequency) [23]. This technique reflects how important a term is to a document, collection, or corpus. It is often used in text mining to define the terms' weighing and relevance.

For each risk, we assume that the larger the number of different keywords of the same risk causes appear in the same report, the stronger is the relationship between that risk and its respective report. In Equation 1, we calculate the *risk report weight* (W_{ij}), where i is the index of a risk and j the index of a report, by dividing the *number of different keywords* (K_{ij}) of a risk's causes and the *maximum number of possible different keywords* ($max(K)$) found on a single report from the whole set of reports.

$$W_{ij} = \frac{K_{ij}}{max(K)} \tag{1}$$

In the same way, we assume that a risk related to more reports has a higher probability (more likely to happen). However, each probability is weighted by the respective risk report's weight (W_{ij}). In Equation 2, we calculate the risk probability (P_i) by summing up all risk reports weight (W_{ij}) and dividing by the *number of reports* (N_i) where the risk causes' keywords were found.

$$P_i = \frac{\sum W_{ij}}{N_i} \tag{2}$$

5. Experiments

This section describes the experimental setup, the technical information about the server and language used in the experiments, and the experimental results highlight the risk probability and features behaviors.

5.1. Experimental Setup

The experiments were conducted on the server (Intel Xeon W-2145 3.7GHz, 4.5GHz Turbo, 8C, 11M Cache, HT, 128GB DDR4, Windows 10 Pro Workstation). We adopted the PHP language with the Laravel framework and MySql Database. Also, for the Crawler, we use the DomCrawler [3] component to manage the connection and data gathering.

5.2. Numerical Results

Based on the content extracted from the reports of ANEEL meetings, several relevant keywords were identified, which are directly associated with 97 of the 233 risks identified by a specialist, presented during the lifecycle of implementation of Transmission Lines (TL).

The association between risks, keywords, and reports made it possible to identify the set of risks in the extracted reports, which areas or phases of the project can be refined. In addition, two equations were generated to be used as risk probability calculation parame-

[3] https://symfony.com/doc/current/components/dom_crawler.html

ters, with the probability of each risk occurring by area or phase of the TL implementation project.

In order to evaluate and visualize the relevance of each keyword, a word cloud was generated (Figure 2). The words displayed (in Portuguese) show a relationship between the ANEEL reports and the risks raised by experts. The greater their frequency, the greater their prominence in the word cloud.

Figure 2. The most relevant keywords identified in the ANEEL reports in Portuguese. Some highlighting terms: *sector agents*, *operation start*, *expected date*, *transmission projects*, *start date*, *commercial operation*.

Figure 3 shows the number of specific risks found in reports from 2005 and 2021. As an example, a risk found was ID 1.11.01, which is described as *Non-approval of the Environmental Study by the licensing agency*. Its causes are pending documentation for the LP request, the elaboration of the environmental study in disagreement with the guidelines defined by the environmental agency, and a record of delay in the LP issuance request deadlines. This risk is related to the Environment area in the Feasibility phase (Auction - LP). It was identified by crossing the following keywords: request for issuance, term of reference, elaboration of the study, and pending at the environmental agency.

The year with the highest number of risks was 2013, with 79 risks. For instance, the risk *Need to change the time of displacement of vehicles for works on roads and access roads* also related to the area of Environment and the Implementation phase (LI - LO). Its primary cause is the verification, by the program execution team, of the need to adjust operating hours or restrict activities related to work in noise-sensitive areas such as schools, daycare centers, hospitals, or communities. The keywords that identified the risk were: program execution, need for adjustments, need for alteration, and access routes.

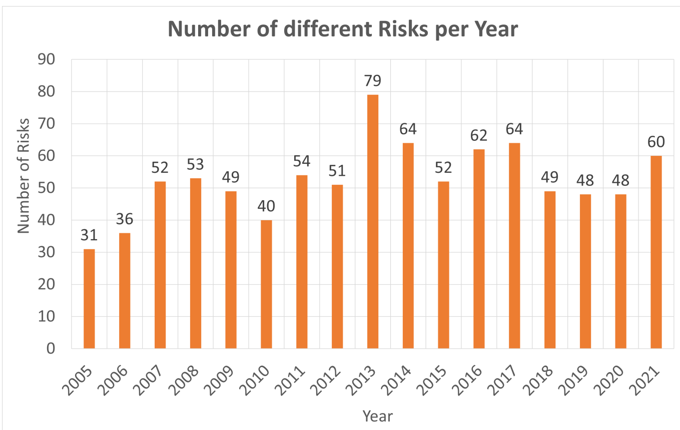

Figure 3. Number of risks per year.

Figure 4 shows the number of distinct risks per report. For example, meetings reports in 2020 were presented and are represented by their internal identifier (ID). The reports with the highest occurrences presented 21 different risks, ID 12.696, held in April in the year in question. It presented risks related to the Environment, Implementation Engineering, and Regulatory areas. The meeting report, ID 12.815, realized in December, contained O&M Engineering and Regulatory risks.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 April 2022 doi:10.20944/preprints202204.0016.v1

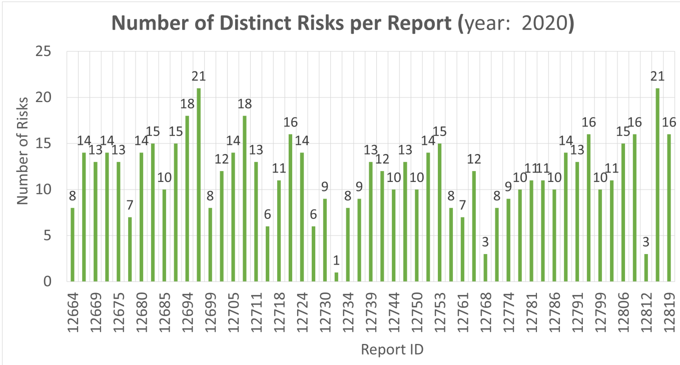

Figure 4. Number of distinct risks per report (year 2020).

Based on Equation 2, the risk probability was calculated as shown in Figure 5. It shows only risks with a probability higher than 20%. The risk identified with the highest probability of occurrence was identifier (ID) 4.04.15, with the following description: *Not meeting the deadlines for replacing equipment already approved by ONS*. This risk is related to the O&M Engineering area and the Operation (O&M) phase; the non-compliance causes it with the ANEEL guidelines in the Electric Sector Accounting Manual and the Electric Sector Asset Control Manual. The main keywords related to risk identification were: accounting for the electricity sector, asset control of the electricity sector, sector accounting manual, manual for asset and asset control.

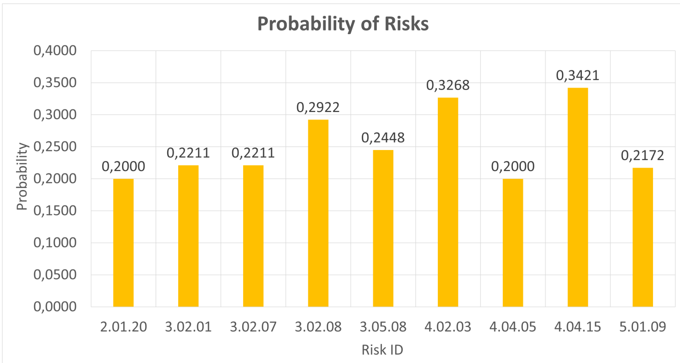

Figure 5. Risk probabilities by each identifier.

Figure 6 presents the risk probability per area of interest: Engineering, O&M Engineering, Land, Environment, Regulatory, and Investors. We can observe that the highest risk probabilities were observed in areas related to engineering, presenting risk keywords such as alteration of the route, fixed assets in service, and charge the costs. Also, the regulatory area contains identified risks by keywords such as changes in the public notice, the anticipation of the commercial operation, and the application of the penalty.

Moreover, Figure 7 shows the risk probability per phase. The phases are Implementation (LILO), Operation (O&M), Pre-Auction, Feasibility (Auction-LP), Feasibility (LP-LI), and Feasibility/Implementation (LP-LI-LO). In this case, the phases that presented the highest risk probability were Implementation (LI-LO) and Feasibility (LP-LI). The Implementation phase presents risk keywords such as opening the administrative process, compliance with conditions, and fixed assets in service. The Feasibility phase contains risks identified by keywords such as plan approval, asset control of the electricity sector, and the existence of pending issues.

Lastly, a web-based *Dashboard* system was developed for the management and visualization of the risk probability. It enables users to interact with graphs and calculate risk probability and impacts on transmission line implementation projects (Figure 8).

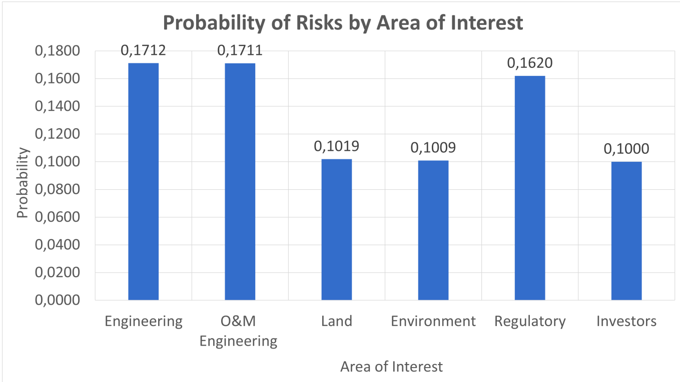

Figure 6. Risk probability per area.

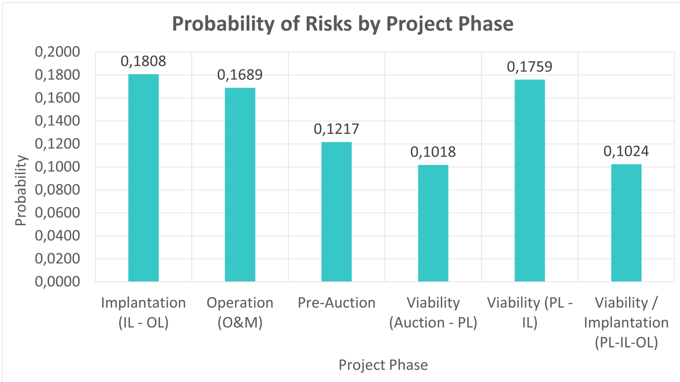

Figure 7. Risk probability per phase.

In addition to the list of risks, Figure 9 shows the risk trend. This graph can be customized by many filters, including time intervals, areas, and phases, to select and generate the results for analysis.

To better monitor managers using the Dashboard, each existing risk will have a monthly evolution graph of its degree and preliminary information regarding the selected risk. For instance, Figure 10 shows the webpage for risk 1.01.03 - Conflict with the Quilombola Remnant Communities (CRQs) regarding the route TL.

6. Conclusions

The application of NLP for unstructured data analysis of the electricity sector made it possible to create an algorithm to calculate the probability estimate of the occurrence of a given risk on the design of a transmission line. As long as there is a document related to the sector, the technique developed can be applied to any data source unstructured text.

Several data sources involve enterprises' transmission lines and decisions related to power lines' electrical transmission. Experts can analyze these sources, and when judged to be relevant, they can be loaded into the algorithm database to calculate the weight of a given document and estimate risk probability.

As future work, it is intended to feed the database with different data sources. From the extraction of its terms, a cross between the different sources, refine the risk weight and probability calculations, and extract information relevant to managers responsible for decision-making.

We also can search for data sources that refer to the imposition of fines on power line projects. This study can analyze the probability of fines, defined as variable installments (PV), depending on the causes of shutdowns in transmission lines. Since every power transmission interruption generates a record and a process at ANEEL for the investigation, the fine is generated if the transmission company (PV) confirms guilty.

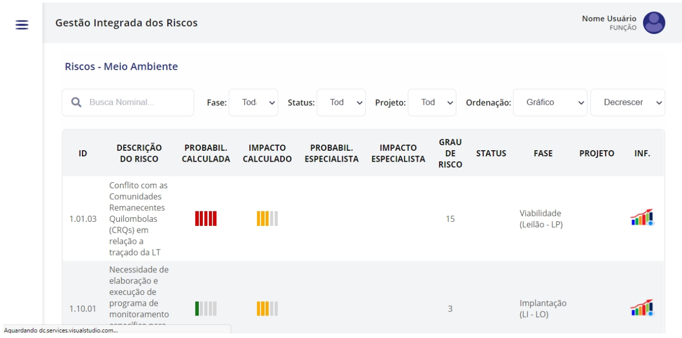

Figure 8. Web-based dashboard system displaying the risks probabilities per area (labels in Portuguese).

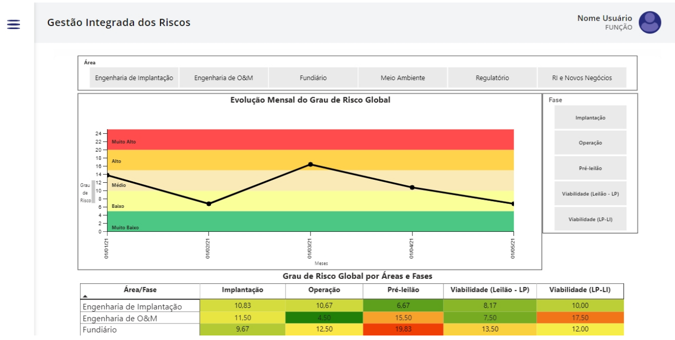

Figure 9. Web-based dashboard system displaying the risk trend (labels in Portuguese).

Author Contributions: L.H.P. and R.B.P designed the project and performed the experiments; P.D.S. verified the analytical methods and supervised the project.; F.G. and F.D.C contributed to the interpretation of the results and wrote the final report; R.S.F. and L.F.L.S. aided in interpreting the results and worked on the analysis of the risks. All authors discussed the results and contributed to the final manuscript.

Funding: This work was supported by the project R&D ANEEL - PD-05456-0001/2019.

Data Availability Statement: All the database used in this work was obtained from the reports of ANEEL meetings on its website http://antigo.aneel.gov.br/pautas-e-atas.

Acknowledgments: The authors gratefully acknowledge financial support from the AVSystemGeo, Equatorial, CNPq, FAPEMIG, CAPES, UFMG, and PUC Minas.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANEEL	Brazilian National Electric Energy Agency
FUNAI	Justice's National Foundation for Indigenous Affairs

References

1. Sadeghi, M.; Shavvalpour, S. Energy risk management and value at risk modeling. *Energy Policy* **2006**, *34*, 3367–3373. doi:https://doi.org/10.1016/j.enpol.2005.07.004.
2. Bhusal, N.; Abdelmalak, M.; Kamruzzaman, M.; Benidris, M. Power System Resilience: Current Practices, Challenges, and Future Directions. *IEEE Access* **2020**, *8*, 18064–18086. doi:10.1109/ACCESS.2020.2968586.
3. Analui, B.; Scaglione, A. A Dynamic Multistage Stochastic Unit Commitment Formulation for Intraday Markets. *IEEE Transactions on Power Systems* **2018**, *33*, 3653–3663. doi:10.1109/TPWRS.2017.2768384.
4. Du, E.; Zhang, N.; Kang, C.; Xia, Q. Scenario Map Based Stochastic Unit Commitment. *IEEE Transactions on Power Systems* **2018**, *33*, 4694–4705. doi:10.1109/TPWRS.2018.2799954.
5. Zhang, N.; Kang, C.; Xia, Q.; Liang, J. Modeling Conditional Forecast Error for Wind Power in Generation Scheduling. *IEEE Transactions on Power Systems* **2014**, *29*, 1316–1324. doi:10.1109/TPWRS.2013.2287766.

 doi:10.20944/preprints202204.0016.v1

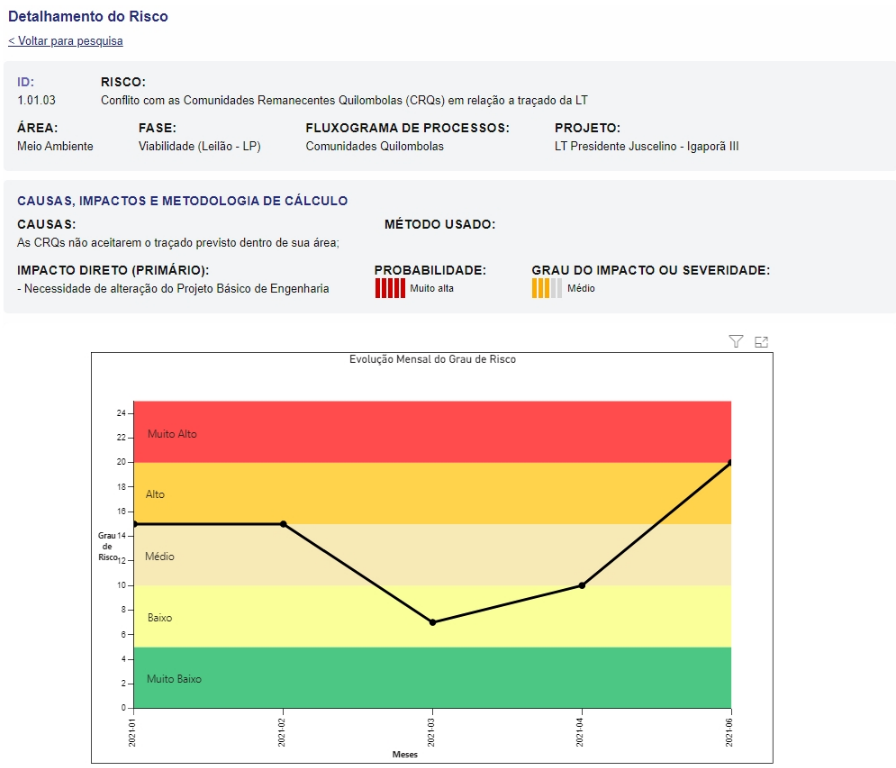

Figure 10. Web-based dashboard system displaying the risk details (labels in Portuguese).

6. Romero, N.R.; Nozick, L.K.; Dobson, I.D.; Xu, N.; Jones, D.A. Transmission and Generation Expansion to Mitigate Seismic Risk. *IEEE Transactions on Power Systems* **2013**, *28*, 3692–3701. doi:10.1109/TPWRS.2013.2265853.
7. Linares, P. Multiple criteria decision making and risk analysis as risk management tools for power systems planning. *IEEE Transactions on Power Systems* **2002**, *17*, 895–900. doi:10.1109/TPWRS.2002.800991.
8. Bruno, S.; Ahmed, S.; Shapiro, A.; Street, A. Risk neutral and risk averse approaches to multistage renewable investment planning under uncertainty. *European Journal of Operational Research* **2016**, *250*, 979–989. doi:10.1016/j.ejor.2015.10.01.
9. Ni, M.; McCalley, J.; Vittal, V.; Tayyib, T. Online risk-based security assessment. *IEEE Power Engineering Review* **2002**, *22*, 59–59.
10. Kirschen, D.; Jayaweera, D. Comparison of risk-based and deterministic security assessments. *IET Generation, Transmission & Distribution* **2007**, *1*, 527–533.
11. Tummala, V.R.; Burchett, J.F. Applying a risk management process (RMP) to manage cost risk for an EHV transmission line project. *International journal of project management* **1999**, *17*, 223–235.
12. Chermack, T.J. *Scenario planning in organizations: how to create, use, and assess scenarios*; Berrett-Koehler Publishers, 2011.
13. Ralston, B.; Wilson, I. *The scenario-planning handbook: a practitioner's guide to developing and using scenarios to direct strategy in today's uncertain times*; South-Western Pub, 2006.
14. Ekel, P.; Pedrycz, W.; Pereira Jr, J. *Multicriteria decision-making under conditions of uncertainty: a fuzzy set perspective*; John Wiley & Sons, 2019.
15. Izadi, M.; Safdarian, A. Financial Risk Evaluation of RCS Deployment in Distribution Systems. *IEEE Systems Journal* **2019**, *13*, 692–701.
16. Liu, K.; Wu, X.; Shi, C. Risk early warning of distribution power system based on data mining technology. 2017 China International Electrical and Energy Conference (CIEEC), 2017, pp. 40–45.
17. Hao, J.; Jinming, C.; Yajuan, G. Data-driven lean Management for Distribution Network. 2018 China International Conference on Electricity Distribution (CICED), 2018, pp. 701–705.
18. Zhao, X.; Luo, L.; Ma, G.; Cai, Z.; Gu, Z.; Wang, Q. Operation and Maintenance Management and Decision Analysis in Distribution Network Based on Big Data Mining. 2018 International Conference on Power System Technology (POWERCON), 2018, pp. 4855–4861.
19. MENEZES, V.P.d. Linhas de transmissão de energia elétrica–aspectos técnicos, orçamentários e construtivos. *Rio de Janeiro* **2015**.
20. Lim, E.P.; Chen, H.; Chen, G. Business Intelligence and Analytics: Research Directions. *ACM Trans. Manage. Inf. Syst.* **2013**, *3*. doi:10.1145/2407740.2407741.
21. Parvizimosaed, M.; Farmani, F.; Monsef, H.; Rahimi-Kian, A. A multi-stage Smart Energy Management System under multiple uncertainties: A data mining approach. *Renewable Energy* **2017**, *102*, 178–189.

22. Ciarapica, F.; Bevilacqua, M.; Antomarioni, S. An approach based on association rules and social network analysis for managing environmental risk: A case study from a process industry. *Process Safety and Environmental Protection* **2019**, *128*, 50 – 64. doi:https://doi.org/10.1016/j.psep.2019.05.037.
23. Leskovec, J.; Rajaraman, A.; Ullman, J.D. *Mining of Massive Datasets*, second ed.; Cambridge University Press, 2014.

doi:10.20944/preprints202204.0016.v1