Methodology to build labeled corpora and classification models to assess technological readiness: a case study with defense technologies described by texts in pt-br

José Voltan[1], Romullo Girardi[1], Thassia Santos[2], Daniel Abreu[3], Júlio Farias[4], Diogo Araujo[1], Víctor Cruz[5], Ronaldo Goldschmidt[1]

[1]Instituto Militar de Engenharia - Rio de Janeiro - Brazil

[2]Universidade Presbiteriana Mackenzie - São Paulo - Brazil

[3]Exército Brasileiro - Brazil

[4]Instituto Tecnológico de Aeronáutica - São José dos Campos - Brazil

[5]Universidade de Brasília - Brasília - Brazil

voltan.jose@ime.eb.br

Abstract. *The Technology Readiness Levels (TRL) scale is a tool for assessing the maturity of technologies. It has been increasingly used by governments and industries to implement tasks such as technological trends detection. The complexity of the assessment process has demanded automated solutions based on classification models that analyze documents to infer the TRL of the technologies described by the texts. However, the lack of corpora with labeled documents has hindered the development of such solutions. To fill this gap, this study proposes a methodology for building corpora labeled w.r.t. the TRL scale. It was applied to a case study and generated a corpus with 168 documents. This corpus was used to develop 30 classification models that hit 51.72% average F1-score.*

1. Introduction

The development of products or critical technologies within the context of a complex project typically involves the elaboration of different subsystems by different teams. In this process, issues may arise, such as communication difficulties among teams regarding the technological development status and reliability of the involved subsystems. Additionally, another issue is the need for efficient resource allocation among the different project components [Mankins 2009] [Girardi et al. 2022].

In this sense, National Aeronautics and Space Administration (NASA) employee Stan Sadin proposed the Technology Readiness Levels (TRL) scale in the 1970s aiming to enable the assessment of the maturity and development of a technology or product. Subsequently, Mankins, another NASA employee, has implemented several modifications, such as increasing the number of levels and providing more precise definitions [Mankins 1995]. Currently, the scale comprises nine levels, as illustrated in Figure 1. As the TRL level increases, the maturity of the technology or product also increases [Mankins 2009].

From the 2000s, the TRL scale began to be adopted by various organizations, governments, and industries, such as the United States Department of Defense and the European Space Agency [Mankins 2009]. In Brazil, several institutions, including the

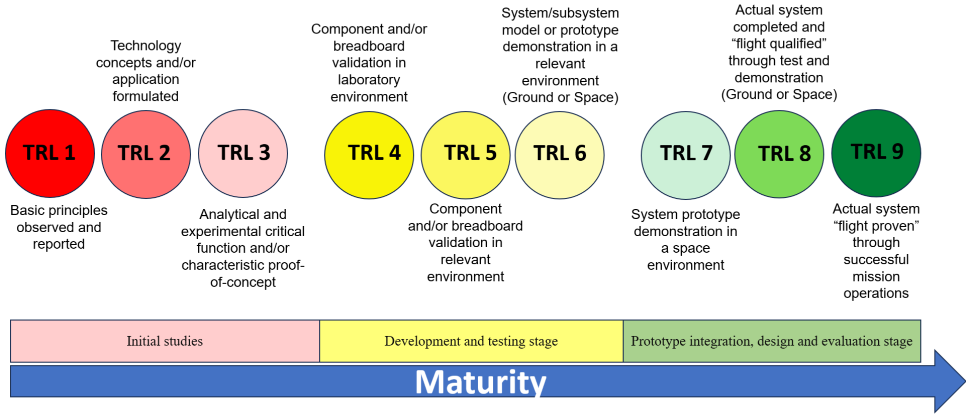

Figure 1. Technology Readiness Levels. Adapted from [Mankins 2009] and [Girardi et al. 2022]

Brazilian Space Agency and the Brazilian Army, have already adopted the TRL scale [Girardi et al. 2022]. The Ministry of Science, Technology, and Innovation has implemented several initiatives for the utilization of the TRL scale, including the incorporation of its concepts into the analysis process of the 'Lei do Bem', an important tool for stimulating R&D investments [Brasil 2020]. These examples of adoption of the TRL scale reinforce the importance that the scale has gained in industry and governments.

In the context of the TRL scale, the process of assessing the maturity of technologies is known as Technology Readiness Assessment (TRA). The TRA is a complex process usually conducted by a team of experts with extensive knowledge of the TRL scale and the technology [Britt et al. 2008]. However, the process with a team of experts is considered expensive, slow, and not scalable, as well as associated with a certain degree of subjectivity [Lezama-Nicolás et al. 2018]. As a way to mitigate these issues, some studies, e.g., [Britt et al. 2008] and [Hardiyati et al. 2018], point to automated solutions, which employ, for example, text mining (e.g., classification models) in the analysis of textual documents that describe the development of technologies. These models analyze the documents to infer the TRL (class) of the technologies described by the texts. Such modeling requires *corpora* where documents must be labeled according to the TRL scale.

In this scenario, the authors of [Silalahi et al. 2018] investigated the classification of scientific publications in Indonesia in the context of biomedicine. To achieve this, they employed a maturity scale with four levels, derived from the TRL scale. However, the *corpus* used in the study was not made available. As far as it was possible to observe, such work illustrates a recurring gap in other scientific reports on the same subject: the unavailability of labeled *corpora* that allow the replication of experiments and comparison among different classification algorithms. Similarly, the mentioned studies do not describe how to label the documents. This lack of *corpora* with documents labeled according to the TRL scale has hindered the development of TRA automated solutions.

To fill this gap, this study proposes a methodology for building *corpora* with documents labeled in the context of the TRL scale. The methodology encompasses everything from document collection to selection and labeling by domain experts. To illustrate its practical feasibility, the results of a case study in TRA comparing different classification algorithms applied to a corpus generated and made available by this methodology are also presented. The corpus generated contains documents written in the Portuguese language (pt-br) and covers technologies within the defense area.

The next sections of this paper were organized as follows: Section 2 presents the proposed methodology for *corpus* construction. In Section 3, this methodology is applied to a case study. Finally, Section 4 presents the concluding remarks highlighting the expected contributions and the direction of the ongoing work.

2. Methodology for Corpus Construction

This section aims to present the proposed methodology for constructing labeled *corpora* in the context of TRA and considerations about document classification according to the TRL scale. The proposed methodology is illustrated graphically in Figure 2. The application of the methodology begins after defining the problem and analysis domain.

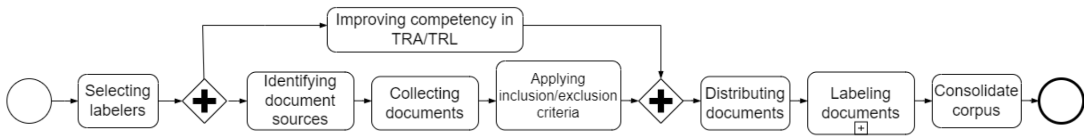

Figure 2. Methodology for Corpus Construction

The first step of the proposed methodology involves selecting a team $R = \{r_1, ..., r_{|R|}\}$[1] of experts in the application domain who will support the document labeling process. Ideally, this group should consist of professionals who not only understand the domain of the documents to be labeled but also know TRA and TRL. However, this dual expertise is challenging to find in a single individual, so selected domain experts who are not familiar with TRA and TRL scale should take part in a training stage.

This training stage comprises videos, meetings, and documents for individual study. In this context, for this work, a series of 10 videos on the topic, totaling 71 minutes, was recorded. These videos are available at https://bit.ly/videoTRL. In addition to explaining concepts about the TRL scale, the videos also introduce the standardization of terms and concepts to be considered by labelers during the labeling process. Besides the videos, this stage involves holding meetings for clarification of doubts, as well as providing articles for study, such as: [Mankins 2009], [Girardi et al. 2022] e [ABNT 2015].

In parallel with the training stage, the proposed methodology includes a step focused on identifying sources of relevant documents. The main sources include databases containing scientific articles, technical reports, test results, requirement specifications, news, patents, and industrial properties. To facilitate the identification of a set of relevant document sources $F = \{f_1, f_2, ..., f_{|F|}\}$ to be considered, interviews with domain experts should be conducted, and the most prevalent sources indicated by the experts should be prioritized. From F, the collection of documents related to the analyzed technology or product is carried out. At the end of this stage, we have a collection $D = \{d_1, ..., d_{|D|}\}$, whose elements need to be labeled by R.

In the labeling process, the documents are distributed among the labelers, as illustrated in Figure 3. This process should take into consideration the affinities between

[1]The notation $|X|$ represents the cardinality, i.e., the number of elements, of an arbitrary set X.

labeler and technology (i.e., academic background and professional experience). During the selection process and meetings held in the previous stages, the labelers must present a set of words related to technologies and products with which they are most familiar.

Thus, for a document d_i, a set of labelers S_i, where $S_i \subset R$. These individuals must independently label the document d_i. For more reliable labeling, it is advisable to adopt $|S_i| \geq 2$. The proposed labeling process is supported by a questionnaire to guide and direct the identification of the most suitable TRL level for the document.

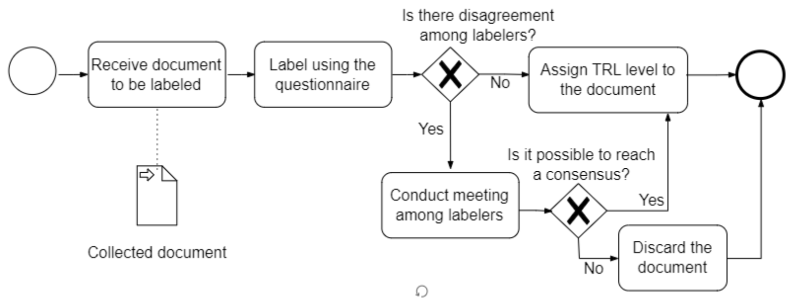

Figure 3. Document labeling

Thus, let $r_k(d_i)$ be the labeling assigned by labeler r_k to document d_i. After the labeling process, each document d_i will have a set of labels $R(d_i) = \{r_j(d_i)/r_j \in S_i\}$. If there is a disagreement in labeling, i.e., there exist $r'(d_i), r''(d_i) \in R(d_i)$ such that $r'(d_i) \neq r''(d_i)$, it should be evaluated whether the labelers can reach a consensus on the TRL level through meetings, or if d_i should be disregarded.

After consolidating the labels, the documents should be stored in JSON (JavaScript Object Notation) files, which offer the advantage of allowing key-value association. Thus, each document should be recorded with the fields ID, title, author, URL (Uniform Resource Locator), document type, label (assigned TRL), abstract, and text. Such organization facilitates the training and evaluation of Machine Learning (ML) models, as well as enabling storage in document-oriented databases, such as MongoDB.

3. Case Study

The proposed methodology was applied to a case study focused on the defense industry domain, restricted to technologies related to computer engineering, electronics, and telecommunications. This scope was chosen because the defense industry is known for producing cutting-edge technologies at the forefront of knowledge with dual applications (i.e., military and civil) [Querino 2022]. E.g.: the internet was originated from the ARPANET (a military network), and the microwave ovens were discovered as a result of experiments during WWII involving radars [Bueno 2022]. The case study was restricted to technologies related to the Brazilian Army, with documents originally written in pt-br.

Concerning the first step of the proposed methodology (selection of labelers), a team of six specialists was chosen: three computer engineers, two electronic engineers, and one telecommunications engineer. All were linked to the defense area and had 5 or more years of experience. To further enhance the team's understanding of TRL/TRA, as well as standardizing some procedures during labeling, a training session was conducted using the videos and articles mentioned in the previous section.

For the identification of relevant public document sources, in addition to the team of labelers, four other domain experts were interviewed. In these interviews, the set of sources $F = \{$*Noticiário do Exército Brasileiro, Simpósio de Aplicações Operacionais em Áreas de Defesa, Revista Militar de Ciência e Tecnologia)*$\}$ was identified. The *Noticiário do Exército Brasileiro* is intended for the publication of news of interest to the Army, such as military operations achievements, receipt of equipment, and conducting tests and trials on products. The *Simpósio de Aplicações Operacionais em Áreas de Defesa* is held annually by the Aeronautics Institute of Technology (ITA). The event has had 25 editions so far with a total audience exceeding 12,000 people [SIGE 2023]. Finally, the *Revista Militar de Ciência e Tecnologia* is produced by the Military Institute of Engineering (IME), being a scientific journal with over 40 years of existence, focused on Science and Technology in the field of Defense [EB Revistas 2023]. All these sources are published in Portuguese.

In this case study, a grouping of TRL was adopted. This simplification is aligned with the approach of technological foresight, as in [Lezama-Nicolás et al. 2018], [Silalahi et al. 2018] e [Hardiyati et al. 2018]. Just as used by [Girardi et al. 2022], three ranges were adopted: Range 1 (TRL 1 to 3), Range 2 (TRL 4 to 6), Range 3 (TRL 7 to 9).

Having made these considerations, after collecting the documents, they were distributed among the labelers, taking into account the affinity between the technology addressed in the documents and each labeler's technical background. After distribution, each labeler read and analyzed each assigned document. To assist in the labeling task, the questionnaire available at https://bit.ly/TRASurvey was used. After answering the questionnaire, the TRL range assigned by the labeler for that document was indicated, or the suggestion for discarding. Each document was analyzed by two labelers (i.e., $|S_i| = 2$). In cases where the two labelers initially did not converge to the same TRL range, meetings were held. This occurred in 14 documents (7.5% of the corpus). From these meetings, either a common range was reached or the decision was made to discard the document.

3.1. Analysis of the corpus

Applying the proposed corpus construction methodology to the present case study, 187 documents have been collected. The documents are distributed as follows: Range 1 (94 documents), Range 2 (41 documents), Range 3 (33 documents), and discarded (19 documents). The constructed corpus is available at https://bit.ly/datasetTRL

The generated *corpus* can be used for training models that combine vector representation techniques with classification models, thus automating the TRA process. In our case study, we used the TF-IDF technique along with six classical ML classification algorithms for classifying documents in the TRL ranges. The preprocessing of each document reduced text dimensionality [Jurafsky and Martin 2022], and cross-validation with k-folds (k=5) was employed for training and evaluation. The codes for this process are available at https://bit.ly/TRLmodels, with Figure 4 illustrating the experimental steps.

Table 1 presents the results of the cross-validation, displaying the average precision, recall, and F1-score (macro). The mean μ presented in the table is followed by the standard deviation σ, i.e., $\mu \pm \sigma$. In bold, the highest mean per performance metric is highlighted, i.e., the algorithm that provided the classification model with the best performance (highest value) for that metric. The last line (*Random*) considered the scenario where the classifier labeled the majority class (range 1) to all instances in its predictions.

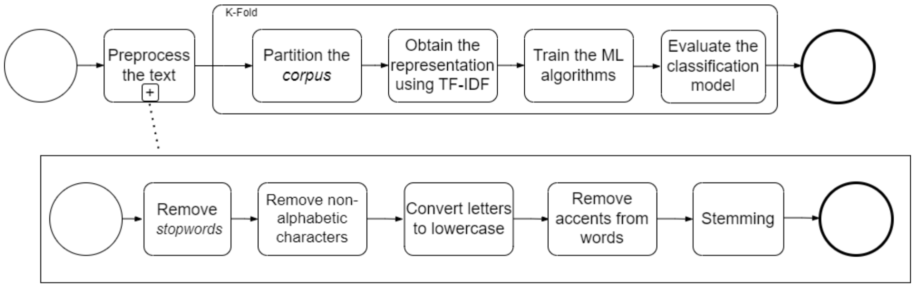

Figure 4. Steps of the proposed experiment

The experimental results indicated that KNN outperformed all other algorithms across all metrics, and all ML algorithms performed better than the *random* baseline. The analysis revealed that the models effectively labeled Range 1 but showed poorer performance in Ranges 2 and 3, likely due to the larger number of documents in Range 1.

Table 1. Results

Algorithm	Precision	Recall	F-1 score
Multinomial Naive Bayes (MNB)	0.502 ± 0.040	0.544 ± 0.046	0.502 ± 0.029
Complement Naive Bayes (CNB)	0.509 ± 0.057	0.608 ± 0.052	0.549 ± 0.053
K-Nearest Neighbors (KNN)	**0.664** ± 0.040	**0.635** ± 0.034	**0.587** ± 0.075
Support Vector Machine (SVM)	0.602 ± 0.158	0.603 ± 0.065	0.489 ± 0.139
Random Forest (RF)	0.614 ± 0.172	0.626 ± 0.070	0.549 ± 0.051
AdaBoost (Adb)	0.593 ± 0.089	0.474 ± 0.149	0.427 ± 0.097
Random	0.187 ± 0.003	0.333 ± 0.000	0.240 ± 0.002

4. Final Considerations

The TRL scale has gained prominence as a way to track the maturity of a technology development project or product, or even in technological forecasting, monitoring the development of new technologies . Traditional solutions for the TRA process are regarded as slow, costly, and not scalable. Consequently, due to these limitations, some authors advocate for automated solutions [Voltan et al. 2024].

An automated approach based on ML typically employs a dataset to train the models. From this perspective, an important gap is the lack of labeled *corpora* and a standardized methodology for *corpora* creation. In face of this gap, this work presented its main contribution: a methodology for creating *corpora* with documents labeled using the TRL scale. This methodology was applied in a specific domain, resulting in a *corpus* consisting of 168 documents in pt-br, with different technologies and products within the scope of defense, in the areas of computer engineering, electronics, and telecommunications.

These promising results indicate that solutions for document classification on the TRL scale may be feasible. Therefore, as future work initiatives, the utilization of other language models for vector representation of documents (e.g., BERT and GPT) and other classification algorithms (e.g., recurrent deep neural networks) are considered. Another research opportunity would be the construction of a larger *corpus*. Additionally, to mitigate the class prevalence problem, LLM (*Large Language Models*) can be used to enrich the produced *corpus* by generating new documents based on the existing ones.

References

ABNT (2015). *NBR ISO 16290:2015: Sistemas espaciais — Definição dos níveis de maturidade da tecnologia (TRL) e de seus critérios de avaliação.*

Brasil (2020). Portaria MCTIC nº 1.892, de 27 de abril de 2020. *Diário Oficial da União*.

Britt, B. L., Berry, M. W., Browne, M., Merrell, M. A., and Kolpack, J. (2008). Document classification techniques for automated technology readiness level analysis. *Journal of the American Society for Information Science and Technology*, 59(4):675–680.

Bueno, C. (2022). Ciência para a guerra e para a paz: uso militar ajudou a ciência a avançar, mas o papel da ciência na busca pela paz é fundamental. *Ciência e Cultura*, 74.

EB Revistas (2023). Sobre a revista— revista militar de ciência e tecnologia. http://www.ebrevistas.eb.mil.br/CT/about. 12 jul. de 2023.

Girardi, R., França, A., and Galdino, J. (2022). A customização de processos de avaliação de prontidão tecnológica baseados na escala TRL: desenvolvimento de uma metodologia para o Exército Brasileiro. *Coleção Meira Mattos*, 16:491–527.

Hardiyati, R., Silalahi, M., Amelia, M., Nadhiroh, I. M., Rahmaida, R., and Handayani, T. (2018). A conceptual model for classification of biomedicine research. *IOP Conference Series: Earth and Environmental Science*, 197(1):012006.

Jurafsky, D. and Martin, J. H. (2022). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. no prelo, USA, 3st edition.

Lezama-Nicolás, R., Rodriguez, M., Rio-Belver, R., and Bildosola, I. (2018). A bibliometric method for assessing technological maturity: the case of additive manufacturing. *Scientometrics*, 117.

Mankins, J. C. (1995). Technology readiness level – a white paper. Technical report, NASA, Washington, D.C.

Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica*, 65(9):1216–1223.

Querino, L. d. F. (2022). *Movimentos societários da indústria de defesa brasileira*, pages 136–168. Editora FGV.

SIGE (2023). SIGE - simpósio de aplicações operacionais em Áreas de defesa. http://https://www.sige.ita.br/. 12 dez. de 2023.

Silalahi, M., Hardiyati, R., Nadhiroh, I. M., Handayani, T., Amelia, M., and Rahmaida, R. (2018). A text classification on the downstreaming potential of biomedicine publications in indonesia. In *2018 ICOIACT*, pages 515–519.

Voltan, J. L. N., Girardi, R., Galdino, J. F., and Goldschmidt, R. R. (2024). Uma revisão e classificação das técnicas de avaliação de prontidão tecnológica baseadas na escala TRL. *Coleção Meira Mattos: Revista das Ciências Militares*, 18(61):49–76.