

Accessible Legislative Information: Automatic Summarisation of Zambian Legislative Documents

1st FName1 LName1

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

fname1.lname1@institution.xx

2nd FName2 LName2

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

fname2.lname2@institution.xx

3rd FName3 LName3

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

fname3.lname3@institution.xx

4th FName4 LName4

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

fname4.lname4@institution.xx

5th FName5 LName5

dept. name of organization (of Aff.)

name of organization (of Aff.)

City, Country

fname5.lname5@institution.xx

Abstract—The National Assembly of Zambia produces a number of important legislative documents which are publicly accessible via its Website. One of the documents published by the National Assembly of Zambia are Bills and Acts, which all form the Laws of Zambia. The Acts and Bills are ideally meant to be open and accessible to the general citizenry, however, prior studies conducted have highlighted the lack of ease of access and difficulties with interpretation of such documents and the two main barriers to enabling open and accessible legal information. This paper presents a potentially viable solution to addressing the challenges with facilitating open and accessible legislative documents by leveraging the use of Natural Language Processing techniques for automatically summarising legislative documents. Specifically, the study was aimed at examining barriers faced by individuals when comprehending legislative documents and, additionally, determining the feasibility of implementing NLP modules capable of generating concise summaries of legislative documents. In order to understand challenges faced when comprehending legal documents, 150 undergraduate students were sampled from the University of Zambia using random sampling. To determine the feasibility of using NLP techniques to provide concise summaries of legislative documents, two (2) NLP models—an abstractive summarisation model and extractive summarisation model. A human evaluation strategy was used to perform a comparative evaluation of the two (2) NLP models, in order to determine the more effective approach. A significant portion—approximately 74.29%—of participants reported ‘Never’ (33.99%) or ‘Rarely’ (40.3%) engaging with legislative documents. In contrast, (25.71%) indicated frequent or very frequent interaction. This distribution underscores a significant gap in familiarity and engagement with Zambian legislative materials among the study participants. In assessing participants’ overall understanding of legal documents, the majority (43.5%) expressed a neutral perception, suggesting that they found these documents neither easy nor hard to understand. The majority of the human evaluators had a preference for the abstractive summarisation model, indicating that its brevity, simplicity, and directness as reasons for their choice. In addition, the results of the abstractive summarisation model were stated as being easier to understand.

Index Terms—document summarisation, legislative documents, natural language processing, zambia.

I. INTRODUCTION

The National Assembly of Zambia is mandated by law to “To execute the legislative, oversight, representative and budgetary functions for enhanced democratic governance” [1]. In the 2022-2026 strategic plan [2], the “Strategic Objective 2.2” aims to “Enhance Public Perceptions of the National Assembly” by making parliament open and accessible to the public and, additionally strengthening ICT platforms for public engagement.

While parliament, and entities such as the Zambia Legal Information Institute (ZambiaLII) [3], publicly makes available important legislation, interpretation of the documents is problematic due to the size of the documents and the vocabulary used. Masson and Tahir report that the barriers associated with providing open and accessible legal information relies on two factors: ease of access and the capacity to interpret the documents [4].

The main objective of the study presented in this paper was to develop an automated summarisation system that utilises the potential of Natural Language Processing (NLP) to greatly enhance the accessibility and understanding of legislative documents in Zambia for the broader population. The specific objectives of the study were as follows:

- To determine the challenges that people have in understanding legal documents
- To design and implement NLP models for summarising legislative documents
- To evaluate the effectiveness of the natural language processing model

This paper is organised as follows: Section I provides context and background information associated with the studies conducted; Section II comprehensively discusses existing work; Section III outlines the methodological approaches employed when conducting the studies; Section IV discusses the

findings and, finally, Section V outlines concluding remarks and potential future work.

II. RELATED WORK

A. Legislative Document Complexities

The specific challenges associated with comprehending legislative documents are discussed by Masson and Tahir, where they place emphasis on interpretation of legal information [4]. The study by Masson and Tahir highlighted the difficulties with accessing legal documents being linked to administrative, financial and technique barriers. In addition, the study identified common sources of legislative documents, such as National Assembly. More significantly, however, one of the findings from the study was the difficulties Civil Society Organisations face interpreting legal and legislative documents.

This paper proposes an automated approach to interpreting and hence summarising legislative documents.

B. Legislative Document Summarisation

There are two classical techniques that are used during document summarisation: abstractive summarisation and extractive summarisation. Abstractive summarisation techniques generate summaries which capture the main information and meaning of textual input using new text, while extractive summarisation functions by identifying important elements from the original text [5].

Numerous studies have been conducted to explore effective techniques for summarising document, with automatic summarisation applied on scientific articles [6], new articles [7] and medical documents [8].

Automatic summarisation has also been applied to legal documents [9] and, more specifically, legislative document; for instance Kornilova and Eidelman [10] benchmark extractive techniques on BillSum, a dataset for summarisation of US Congressional and California state bills with their results demonstrating the potential of transfer learning.

The work presented in this paper focuses on legislative documents prepared and formatted by the National Assembly of Zambia, presenting unique challenges specific to this context.

C. Document Summarisation Evaluation

Evaluation of document summarisation primarily uses two approaches: automatic summarisation techniques and human evaluation [11]. Automatic evaluation largely uses traditional metrics such as ROUGE¹ or BERTScore², while human evaluation approaches largely focus on subjective quality of resulting text.

The work presented in the paper employed human evaluation techniques as the primary goal was to assess the comprehensiveness of summaries, comparatively.

III. METHODOLOGY

A mix-methods approach was employed to conduct the studies outlined in this paper as follows:

- A survey, outlined in Section III-A was conducted with students at University X
- NLP summarisation models were implemented as outlined in Section III-B
- A controlled comparative evaluation was performed with human evaluators, as outlined in Section III-C, in order to determine the perceived effective summarisation technique

A. Understanding Legislative Documents

In order to understand participants' with legislative documents and, additionally, challenges associated with comprehending legislative documents, a survey was conducted with randomly sampled full-time undergraduate students at The University of Zambia. Sociodemographic factors—gender, age, school/faculty, programme of study and year of study—were collected from participants, in addition frequency with which legislative documents were accessed and, finally, difficulties with comprehending legislative documents using "Statutory Instrument No. 12 of 2018"—shown in Figure 1—as a reference.

B. Summarisation Models: Design and Implementation

The implementation of the NLP summarisation models and pipelines was guided by the Cross-industry Standard Process for Data Mining (CRISP-DM) model [12], with the six (6) phases used as follows:

- Phase 1.** Business Understanding—Acts and Bills available on the National Assembly of Zambia Website were analysed to understand the structure of legislative documents
- Phase 2.** Data Understanding—Data sources were identified, with data extraction conducted on publicly available PDF documents
- Phase 3.** Data Preparation—Data was pre-processed using common text pre-processing techniques
- Phase 4.** Modelling—NLP models were implemented in order to summarise legislative documents
- Phase 5.** Evaluation—Human evaluators were used to determine the most effective summarisation technique—abstractive and extractive summarisation
- Phase 6.** Deployment—A simple Web-based interface was implemented to comparatively evaluate the two abstractive and extractive summarisations models.

In order to implement effective NLP summarisation models, focus group discussions were held with University of Zambia Law students, in order to understand elements that comprise legislative documents.

Two NLP summarisation models were implemented use two common summarisation techniques: abstractive summarisation and extractive summarisation.

¹<https://huggingface.co/spaces/evaluate-metric/rouge>

²<https://huggingface.co/spaces/evaluate-metric/bertscore>

GOVERNMENT OF ZAMBIA		APPENDIX FIRST SCHEDULE (Paragraph 2)	
STATUTORY INSTRUMENT NO. 12 OF 2018		Fee Units	
The Local Government Act (Laws, Volume 16 Cap. 281)			
The Local Government (Street Vending and Nuisances) (Amendment) (No. 2) Regulations, 2018			
IN EXERCISE of the powers contained in section 84 of the Local Government Act, the following regulations are made:			
1. (1) These Regulations may be cited as the Local Government (Street Vending and Nuisances) (Amendment) (No. 2) Regulations, 2018 and shall be read as one with the Local Government (Street Vending and Nuisances) Regulations, 1992, in these Regulations referred to as the principal Regulations.	Title	1. Spitting or vomiting on, or along, a street or prescribed road	333.33
	S. I. No. 134 of 1992	2. Throwing litter on, or along, a street or prescribed road	1,666.67
		3. Urinating in any unauthorised place	2,500.00
		4. Defecating in any unauthorised place	2,500.00
		5. Singing an obscene song or saying obscene words in a street or public place	2,500.00
		6. Writing an obscene word or drawing an indecent figure or representation or, defacing a permanent structure	2,500.00
		7. Wilfully or negligently extinguishing or damaging a street lamp or damaging a lamp-post, telephone or electric light wire, cable insulator or bucket or standard supporting any such cable or, causing an interruption in the supply of electric current by any means whatever, without the permission of the Council or its authorised agent	2,500.00
3. The principal Regulations are amended by the deletion of the First Schedule and the substitution thereof of the Schedule set out in the Appendix	Amendment of First Schedule	8. Plying trade by any licensed hawker within an area more than five days in a calendar month, in the same place for more than thirty consecutive minutes or plying on two or more	

Fig. 1. Statutory Instrument 18 of 2018: Street Vending and Nuisances.

C. Summarisation Models: Comparative Analysis

In order to comparatively evaluate the abstractive and extractive summarisation models, a controlled experiment was conducted with human evaluators.

1) *Task Design*: The National Pension Scheme Act No. 1 of 2023³ was used as input to the two summarisation models and corresponding summaries generated.

2) *Experimental Design*: A comparative analysis between the abstractive summarisation model and extractive summarisation model was performed. The study was conducted using a within-subject design, with counterbalancing applied by altering the summarisation models they initially read and rated.

3) *Procedure*: Each participant was required to sign a consent form and subsequently required to read the "National Pension Scheme Act No. 1 of 2023" document and, additionally, the two corresponding summaries. Finally, participants were required to complete a questionnaire design to evaluate the following:

- Relevance of each of the two summaries, relative to the original document
- Readability of each of the two summaries
- Preferred summary, when comparing the two generated summaries

IV. RESULTS AND DISCUSSION

A. Understanding Legislative Documents

The survey gathered responses from 126 undergraduate students, primarily falling within the 18-25 age group, with a mix of participants from various schools and programs of study. Notably, Humanities and Social Sciences and Veterinary Medicine constituted the majority of participants. The survey covered students from different study years, with a substantial representation from Year 2 and Year 3.

The results indicate that a significant portion of participants (98.4%) have interacted with Zambian legislative documents,

³<https://www.parliament.gov.zm/node/11020>

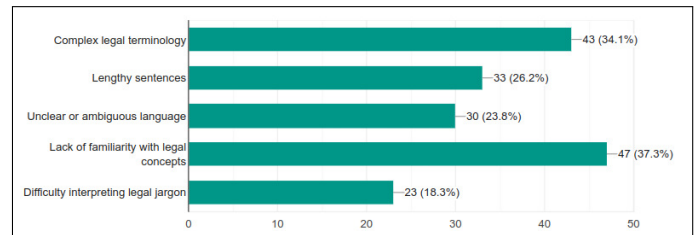


Fig. 2. Summary 2: .



Fig. 3. Summary 2: .

though the frequency varied. The participants provided diverse responses regarding their overall understanding of legal documents, with the majority falling within the mid-range on a scale of 1-5.

Participants highlighted various challenges in understanding legal documents, as shown in Figure 2, including complex legal terminology, lengthy sentences, unclear or ambiguous language, lack of familiarity with legal concepts, and difficulty interpreting legal jargon. These challenges point to potential barriers that may hinder effective comprehension of legislative materials.

Factors such as complex legal terminology, lengthy sentences, unclear language, lack of familiarity with legal concepts, and difficulty interpreting legal jargon were identified as influencing participants' understanding of legal documents, as shown in Figure 3. These insights underscore the need for

targeted interventions to address these specific challenges

Participants provided valuable comments and suggestions, emphasising the importance of brevity and clarity in legal documents. Recommendations included minimising lengthy sentences, avoiding technical jargon, and making legal documents more accessible.

The survey conducted at the University of Zambia sheds light on the challenges undergraduate students face in understanding legal documents. The findings indicate a diverse range of difficulties, encompassing language complexity, sentence structure, and a lack of familiarity with legal concepts. The insights from this study pave the way for potential improvements in legal document accessibility and comprehension. The use of automatic summarization software is suggested as a possible solution to address these challenges, providing concise and clear summaries of legislative materials.

B. Automatic Summarisation

1) *Data Extraction and Pre-processing*: The Python programming language⁴ was used to implement the summarisation models. A two-step process was used to extract data from PDF documents for the purpose of preparing it for subsequent data cleaning. The primary library used for this extraction was PyMuPDF⁵, specifically the fitz module. This library allows for efficient handling of PDF documents in Python.

In order to use a standardised representation of textual input, case folding was applied to input text.

2) *Extractive Summarisation Model*: The evolution of word embedding models progressed with Mikolov et al., who introduced the word2vec model based on skip-grams. This model extended our understanding of word relationships by considering complex associations, including opposites, tenses, plurals, and phrases. Here, the SpaCy⁶ library plays a pivotal role in text processing, aiding in the identification and removal of stop words and punctuation.

Figure 4 illustrates a sample output for the abstractive summarisation model.

3) *Abstractive Summarisation Model*: The abstractive summarisation model was implemented using the BART model from the Hugging Face Transformers library. The code utilized the "transformers" library to import the necessary components, including the pipeline for text generation, the BART tokenizer, and the BART model for conditional generation. Within the "abstractive_summarize" function, the pre-trained BART model and tokenizer were loaded using the specified model name "facebook/bart-large-cnn". The input text was then tokenized, and the summary was generated using the BART model. The "generate" method of the model was utilized to produce the summary, with parameters such as maximum and minimum length, length penalty, number of beams, and early stopping specified to control the summarization process. This approach proved to be successful in producing abstractive

text .extracted from pdf national pension scheme amendment no 1 of 2023 1 an act to amend the national pension scheme act 18th april 2023 enacted by the parliament of zambia 1 this act may be cited as the national pension scheme amendment act 2023 and shall be read as one with the national pension scheme act in this act referred to as the principal act 2 section 2 of the principal act is amended by the insertion of the following new definitions in the appropriate places in alphabetical order preretirement lumpsum benefit means a oneoff benefit payable to a member who has not yet attained pensionable age residual contributions means the remainder of the contributions on a members account after a member has accessed the oneoff preretirement lumpsum benefit and zambia statistics agency means the zambia statistics agency established under the statistics act 2018 3 section 9 of the principal act is amended by the a insertion of the following new subsection immediately after subsection 1 2 despite subsection 1 the authority may pay a oneoff preretirement lumpsum benefit to a member who has met the requirements specified under section 21a and enactment short title amendment of section 2 act no 13 of 2018 cap 256 amendment of section 9 government of zambia act no 1 of 2023 date of assent 14th april 2023 single copies of this act may be obtained from the government printer po box 30136 10101 lusaka price k800 each b renumbering of subsections 2 3 and 4 as subsections 3 4 and 5 respectively 4 section 19 of the principal act is amended by the deletion of subsection 5 and the substitution therefor of the following 5 the authority shall determine the national average earnings under subsection 3 annually by applying statistics and data compiled by the zambia statistics agency 5 the principal act is amended by the insertion of the following new section immediately after section 19 19a the minimum retirement pension of members that access a preretirement lumpsum benefit shall be eighty percent of the minimum monthly pension 6 the principal act is amended by the insertion of the following new section immediately after section 21 21a 1 despite section 21 a member is entitled to a preretirement lumpsum benefit where that member consents in writing to access a preretirement benefit is below pensionable age and has a made a minimum of sixty monthly contributions or b attained the age of forty five 2 the preretirement lumpsum benefit payable under subsection 1 shall be twenty percent of the indexed monthly contributions and the accrued interest 3 the payable benefit at retirement or any other benefit to a member who has accessed the preretirement lumpsum benefit under subsection 1 shall be based on a members residual contributions and contributions made after the receipt of a pre-retirement lumpsum benefit 4 a member who accesses a preretirement benefit under subsection 1 consents to a reduction in that members retirement or other final benefit payable under this act 2 no 1 of 2023 national pension scheme amendment amendment of section 19 insertion of section 19a]

Fig. 4. Summary 1: Extractive Summarisation.

National pension scheme amendment no 1 of 2023 is an act to amend the national pension scheme act 18th april 2023 enacted by the parliament of zambia. The principal act of the act is amended by the insertion of the following new definitions in the appropriate places in alphabetical order preretirement lumpsum benefit means a oneoff benefit payable to a member who has not yet attained pensionable age residual contributions means the remainder of the contributions on a members account after a member has accessed the oneoff preretirements benefit. The authority shall determine the national average earnings under subsection 3 annually by applying statistics and data compiled by the Zambia statistics agency. The minimum retirement pension of members that access a preret Retirement benefit shall be eighty percent of the minimum monthly pension referred to under section 243. A member who suffers an invalidity after accessing a pre-Retirement lump sum benefit is entitled to eighty per cent of the applicable minimum monthly invalidity pension. Despite subsection 1 the authority may pay a one off preret retirement lumpsum benefit to a members who has met the requirements specified under section 21a and enactment short title amendment of section 2 act no 13 of 2018 cap 256 amendment of sections 9 government of 2 Zambia act no 1 2023 date of assent 14th aprils 2023 single copies of this act may be obtained from the government printer po box 30136 10101 lusaka price k800 each b renumbering of subsections 2 3 and 4 as subsections 3 4 and 5 respectively 4 section 19 of the principal act is amend by the deletion of subsection 5

Fig. 5. Summary 2: Abstractive Summarisation.

summaries with the BART model, leveraging its capabilities for text generation and summarization.

Figure 5 illustrates a sample output for the abstractive summarisation model.

By integrating both extractive and abstractive techniques, this project aimed to provide a comprehensive understanding of Zambian legislative materials. The use of diverse Python libraries, including SpaCy for extractive summarization and Transformers BART Model for abstractive summarization, showcased a multifaceted approach to effectively distill and present critical information from lengthy legal documents.

C. Human Comparative Evaluation

The purpose of this evaluation was to assess the effectiveness of automatic summarization techniques applied to Zambian legislative documents. The participants were asked to read two summaries (Summary 1 - extractive summary, and

⁴<https://www.python.org>

⁵<https://github.com/pymupdf/PyMuPDF>

⁶<https://spacy.io>

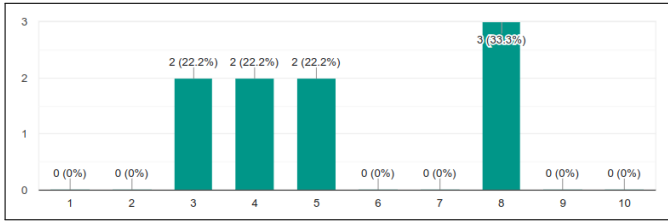


Fig. 6. Summary 1: Relevance.

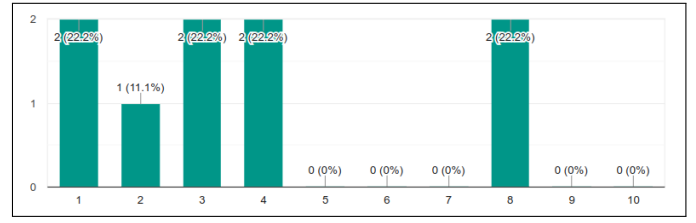


Fig. 8. Summary 1: Readability.

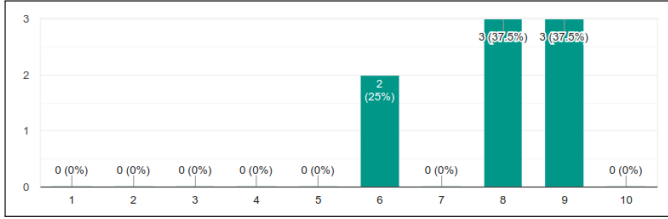


Fig. 7. Summary 2: Relevance.

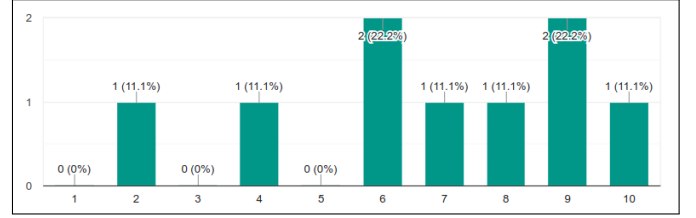


Fig. 9. Summary 2: Readability.

Summary 2 - abstractive summary) generated from the original legislative document. The evaluation focused on relevance, readability, and preference for conveying information.

A total of nine participants participated in the comparative evaluation.

The majority of participants who preferred Summary 2 cited its brevity, simplicity, and directness as reasons for their choice. Participants mentioned that Summary 2 was easier to understand and more straightforward. One participant specifically noted that Summary 2 was "straight to the point," enhancing quick comprehension. Two participants provided feedback, with one expressing a positive sentiment, stating, "This is good." The other participant did not provide specific comments.

The evaluation indicates a strong preference for Summary 2, the abstractive summary, among participants. While both summaries received similar ratings for relevance and readability, the brevity and directness of Summary 2 seemed to resonate more with participants. The positive feedback aligns with the project's goal of enhancing the accessibility of Zambian legislative information through automatic summarization.

In the relevance evaluation, for the extractive summary (Summary 1), participants rated it between 3 and 8, with none selecting 1, 2, 9, or 10. Similarly, for the abstractive summary (Summary 2), ratings ranged from 3 to 10, with no participants choosing 1 or 2. In terms of readability, both summaries achieved average scores around 4.5, reflecting moderate levels of readability. Significantly, the majority of participants (88.9%) expressed a preference for Summary 2, citing its brevity, simplicity, and directness as reasons for their choice. This preference aligns with the feedback received, where participants highlighted that Summary 2 was easier to understand and more straightforward, with one participant specifically noting that it was "straight to the point," facilitating quick comprehension. Out of the two participants who provided feedback, one expressed a positive sentiment,

stating, "This is good," while the other did not provide specific comments.

V. CONCLUSIONS AND FUTURE WORK

The imperative for automating the summarization of Zambian Legislative Documents becomes evident as it presents an opportunity for citizens to gain quick and enhanced comprehension of lengthy legal documents, including Acts of Parliament. This study aimed at identifying challenges in understanding legislative documents, developing and implementing an NLP model for automatic summarization, and assessing the model's effectiveness. The evaluation underscores the efficiency of abstractive summarization techniques, particularly noting participant preference in conveying information. These findings offer valuable insights into the strengths and preferences of various summarization approaches, guiding future initiatives to improve the accessibility and understanding of Zambian legislative documents.

REFERENCES

- [1] National Assembly of Zambia, "About National Assembly," [online] <https://www.parliament.gov.zm/node/4529>, 2023, (Accessed September 23, 2024).
- [2] —, "The National Assembly of Zambia Strategic Plan 2022-2026," [online] <https://www.parliament.gov.zm/sites/default/files/documents/pages/Strategic%20Plan%202026.pdf>, 2022, (Accessed September 23, 2024).
- [3] Zambia Legal Information Institute, "Welcome to the Zambia Legal Information Institute," [online] <https://zambialii.org>, 2024, (Accessed September 23, 2024).
- [4] M. J. Masson and O. Tahir, "The legal information needs of civil society in zambia," *Journal of Open Access to Law*, vol. 4, no. 1, 2016.
- [5] P. Mehta, "From extractive to abstractive summarization: A journey," in *ACL (Student Research Workshop)*. Springer, 2016, pp. 100–106.
- [6] N. I. Altmami and M. E. B. Menai, "Automatic summarization of scientific articles: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1011–1028, 2022.
- [7] H. Oliveira and R. D. Lins, "Assessing abstractive and extractive methods for automatic news summarization," in *Proceedings of the ACM Symposium on Document Engineering 2024*, 2024, pp. 1–10.
- [8] A. Sarker, D. Mollá, and C. Paris, "Extractive summarisation of medical documents using domain knowledge and corpus statistics," *The Australasian medical journal*, vol. 5, no. 9, p. 478, 2012.

- [9] B. Hachey and C. Grover, "Extractive summarisation of legal texts," *Artificial Intelligence and Law*, vol. 14, pp. 305–345, 2006.
- [10] A. Kornilova and V. Eidelman, "Billsum: A corpus for automatic summarization of us legislation," *arXiv preprint arXiv:1910.00523*, 2019.
- [11] J. Steinberger and K. Ježek, "Evaluation measures for text summarization," *Computing and Informatics*, vol. 28, no. 2, pp. 251–275, 2009.
- [12] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000, pp. 29–39. [Online]. Available: <http://cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>