



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Kannda Text Summarization using AI and Machine Learning Algorithms

Sadiya Mulla

Department of Computer Science, Rani Channamma University, Belagavi, India

ABSTRACT

Presenting a Kannada Text Summariser that uses BART (Bidirectional and Auto-Regressive Transformers) to provide brief summaries of Kannada text is the goal of this project. The Django web application was coupled with a Python 3.13 development system to enable users to input lengthy Kannada paragraphs and obtain a concise, well-organised summary. Efficient text summarisation for Kannada material is achieved by integrating sophisticated natural language processing algorithms with an easy-to-use interface. This is an especially helpful tool for users who need to quickly extract meaningful information from large amounts of Kannada textual material.

Keywords: Kannada textual material, RNN, BART (Bidirectional and Auto-Regressive Transformer), Spacy, NLP algorithm, Pytorch

Introduction

The goal of this research is to create a sophisticated AI-based system that can use state-of-the-art Transformer models to summarise Kannada text. Tools that can effectively reduce extensive text into succinct summaries are becoming more and more necessary as the use of digital information in regional languages like Kannada continues to expand. This research attempts to fill that gap by developing a model that can distil a section of Kannada text, usually ten lines long, into a two-line synopsis while retaining all of the important context and meaning.

Utilising Transformer architectures—specifically, BART (Bidirectional and Auto-Regressive Transformers) models—the method outperforms other approaches in problems involving natural language processing. In order to provide excellent summaries, these models will be properly adjusted for the Kannada language, paying close attention to capturing the subtleties of the language's semantics and structure.

The project will start with gathering data, creating an extensive dataset of Kannada text along with matching summaries. To make sure the data is in the best state for training the model, this dataset will be painstakingly cleaned and pre-processed, including operations like tokenisation and the elimination of any unnecessary characters.

Once the dataset is ready, the selected Transformer model will be fine-tuned. This process involves adjusting the model's parameters to adapt it to the specifics of Kannada text summarization, training it to understand how to condense information while retaining the core message. The model's performance will be evaluated using both automated metrics, like ROUGE and BLEU, which measure the quality of the summaries by comparing them to reference summaries, and manual evaluation to ensure that the model-generated summaries are accurate, fluent, and meaningful.

After achieving satisfactory performance, the model will be deployed as a user-friendly web application. This application will allow users to input Kannada text and receive an instant summary, making it a valuable tool for anyone needing to quickly digest large amounts of information in Kannada. The deployment will focus on ensuring that the model is efficient and responsive, capable of handling real-time summarization tasks.

This project not only demonstrates the applicability of advanced NLP techniques to regional languages but also contributes to the development of tools that support the growing digital content ecosystem in Kannada. By providing an effective solution for summarizing Kannada text, this project has the potential to benefit educators, students, journalists, and content creators who work with Kannada language content.

Literature Survey

The literature on Kannada text summarization has evolved significantly, focusing on various techniques and methodologies to enhance the effectiveness of summarization. Geetha and Deepamala (2015) [1] explored the application of latent semantic analysis for Kannada text summarization, highlighting its potential to improve information retrieval. In a more recent study, Pallavi and Sarvamangala (2022) [2] implemented an extractive technique to summarize Kannada text, emphasizing the importance of selection algorithms in distilling key information from documents. Chandrika and Kallimani

(2022) [3] also contributed to the field by employing a page ranking technique for extractive summarization, demonstrating its effectiveness in prioritizing relevant content within Kannada text documents. Jayashree et al. (2012) [5] introduced a categorized sentence ranking method, providing a structured approach to summarize text documents based on thematic relevance. Additionally, Ananth et al. (2022) [6] investigated the nuances of extractive summarization specifically for Kannada, further enriching the body of research. Badiger et al. (2024) [8] presented a comprehensive overview of Kannada text summarization techniques, while Ranganatha et al. (2014) [9] investigated a probabilistic approach for federated document summarization. Finally, Ananth et al. (2021) [10] highlighted the role of parts of speech tagging in improving extractive summarization techniques for Kannada documents, underscoring the interplay between linguistic analysis and summarization. Collectively, these studies reflect a growing interest in developing robust methods for Kannada text summarization, utilizing both traditional and modern machine learning techniques to enhance the quality and efficiency of information extraction.

Methodology

proposed methodology for Kannada text summarization using AI and machine learning techniques. This approach combines both traditional machine learning methods and modern deep learning models for text summarization. Kannada text summarization using AI and machine learning involves two main approaches: **extractive** and **abstractive summarization**. **Extractive summarization** selects key sentences from the text. Techniques like **Text Rank** and machine learning models (e.g., SVM) are used to identify important sentences based on linguistic features such as word frequency or sentence position. **Abstractive summarization** generates new sentences by rephrasing the original content. Deep learning models like **RNNs**, **LSTMs**, **GRUs**, and **Transformer-based models** (such as **BART** and **T5**) are commonly used for this task. These models can be fine-tuned on Kannada-specific data to improve performance. The out of the developed methodology is Kannada text summarization using Abstractive summarization file is as shown in the below figure 1.

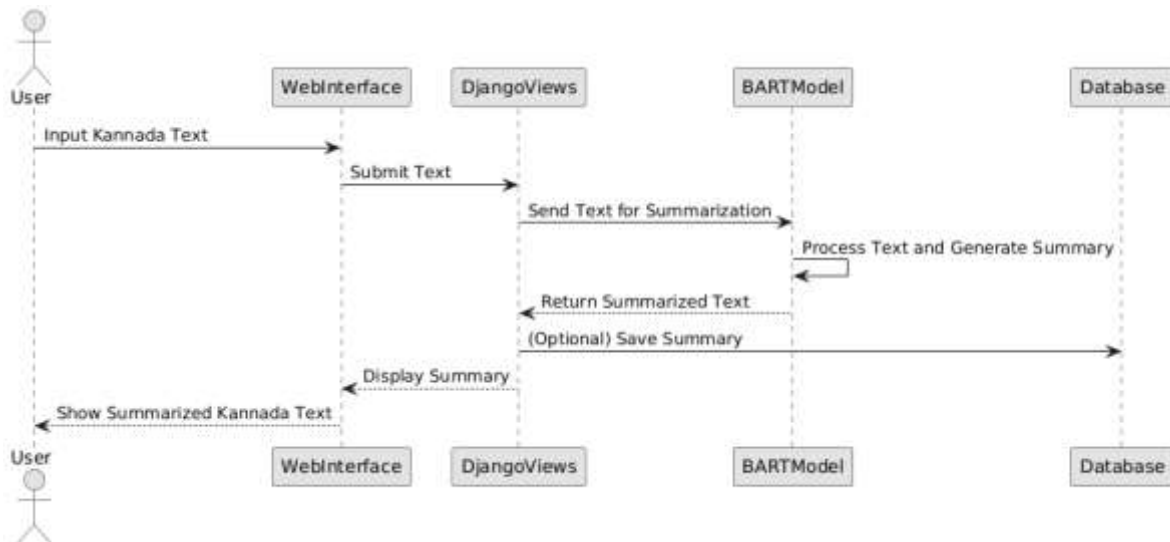


Fig1: The Block diagram Kannada Text summarization using AI and ML

BART (Bidirectional and Auto-Regressive Transformers) is a highly effective model for Kannada text summarization, especially in generating **abstractive summaries**. Unlike extractive methods, which pick key sentences directly from the text, BART creates new sentences that capture the essence of the original content. It follows a **transformer-based encoder-decoder architecture**, where the encoder processes the entire Kannada text, and the decoder generates a concise summary. BART is pre-trained on large datasets, and although Kannada data might be limited, fine-tuning can adapt the model to understand Kannada grammar, syntax, and linguistic nuances. Using **transfer learning** through models like **mBART**, which supports multiple languages, can also be highly beneficial. BART's ability to grasp complex sentence structures, such as compound sentences and formal constructs in Kannada, allows it to produce fluent, contextually accurate summaries. This makes it a powerful tool for Kannada text summarization tasks, offering more coherence and relevance compared to traditional methods. The use of BART in Kannada Text summarization as shown in below figure 2.

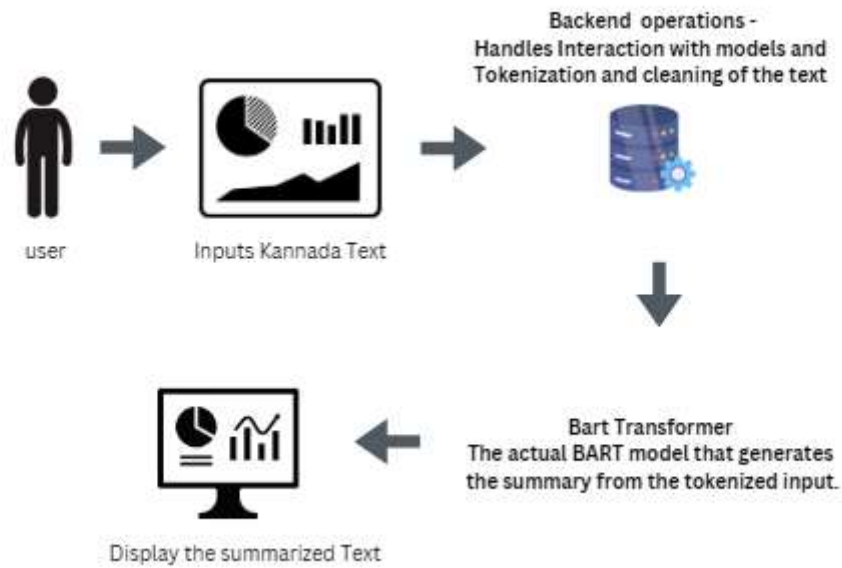


Fig2: FOS framework of Kannada text summarizer

Hardware Requirements

Every computer operating system is made for a specific computer architecture, which is primarily the framework for architecture. A crucial system requirement for any software is the power of the central processor unit (CPU). Additionally, memory, secondary storage, a display adapter, and peripheral objectives are hardware criteria that all models must fulfil. The Kannada text Summarizer needs following hardware:

Hard Disk: 1.2GB or above

RAM: 64MB

Processor: Intel Core i5 or AMD RYZEN 5

Software Requirements

The software requirements are description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. The requirements can be obvious or hidden, known or unknown, expected or unexpected from client's point of view. The Python web application development process can be accelerated with the help of the free and open-source Django web framework. Based on Django frameworks, the Developed Summariser tool is implemented. The main reason for choosing this platform, Django, was to simplify the process of developing and building applications faster.

PROGRAMMING LANGUAGE:

Front End: Python, Django

Back End: HML, CSS, JavaScript

Libraries: TensorFlow, pytorch, Pandas, spacy

Development Environment: Visual Studio cod, cmd

Result and Discussion

A login page for a Kannada text summarizer provides a simple and secure way for users to access the tool. Key features include user authentication with username/email and password, and options for social logins like Google. The interface is fully localized with Kannada text for fields such as username (ಬಳಕೆದಾರ ಹೆಸರು), password (ಪಾಸ್‌ವರ್ಡ್), and login (ಲಾಗಿನ್ ಮಾಡಿ). It supports user registration (ನವೀಕರಿಸಿ) and password recovery (ಪಾಸ್‌ವರ್ಡ್ ಮರೆತಿದಿರಾ?). Security features like captcha and multi-factor authentication (MFA) enhance safety. The page is mobile-friendly and offers real-time error feedback in Kannada for user convenience.



Fig3: Screenshot of login page of Kannada text summarizer

The register page for a Kannada text summarizer allows new users to create accounts easily. It features user information fields like name, email, and password, with full Kannada language localization for accessibility. The page includes password guidelines, captcha verification, and asks users to agree to terms and privacy policies. Users can also sign up via social media. It's mobile-friendly, provides real-time feedback for errors like mismatched passwords, and redirects to the login page after successful registration. Security measures like encrypted passwords and verification emails ensure data safety.



Fig4: Screenshot of Register page of Kannada text summarizer



Fig5: Result of Kannada text summarizer

Conclusion:

The Kannada text summarization system employs advanced Transformer models, such as BART, to generate concise summaries from lengthy Kannada texts. Integrated with a Django web application, it offers an intuitive interface for users to input text and receive summaries, while utilizing SQLite for efficient data management. The system's performance is assessed through automated metrics and user feedback, fostering continuous improvement. This project exemplifies the practical application of cutting-edge NLP techniques in regional languages, enhancing language understanding and information accessibility, and paving the way for future advancements in text summarization and NLP.

Reference:

- [1] Geetha, J. K., & Deepamala, N. (2015, August). Kannada text summarization using latent semantic analysis. In *2015 International conference on advances in computing, communications and informatics (ICACCI)* (pp. 1508-1512). IEEE.
- [2] Pallavi, P., & Sarvamangala, D. R. (2022, July). Kannada Text Summarization using Extractive Technique. In *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1-4). IEEE.
- [3] Chandrika, C. P., & Kallimani, J. S. (2022). Extractive text summarization of kannada text documents using page ranking technique. In *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2021* (pp. 683-695). Singapore: Springer Nature Singapore.
- [4] Chandrika, C. P., & Kallimani, J. S. (2022). Extractive text summarization of kannada text documents using page ranking technique. In *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2021* (pp. 683-695). Singapore: Springer Nature Singapore.
- [5] Jayashree, R., Srikanta, M. K., & Anami, B. S. (2012, November). Categorized text document summarization in the Kannada language by sentence ranking. In *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 776-781). IEEE.
- [6] Veena, R., Ramesh, D., & Hanumanthappa, M. (2024, January). Text Summarization for Kannada Text Documents: A Review. In *International Conference on Universal Threats in Expert Applications and Solutions* (pp. 41-53). Singapore: Springer Nature Singapore.
- [7] Ananth, A., Venugopala, P. S., & Bhat, S. S. (2022, June). An Investigation on the Extractive Summarization of Kannada Text. In *International Conference on Signal & Data Processing* (pp. 315-323). Singapore: Springer Nature Singapore.
- [8] Badiger, K., Sonagaj, S., Giraddi, S., & Reshmi, B. M. (2024, January). Kannada Text Summarization. In *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)* (pp. 1-5). IEEE.
- [9] Ranganatha, S., Vinay, S. K., & Bhargava, H. S. (2014). Federated Document Summarization Using Probabilistic Approach for Kannada Language. *International Journal of Innovative Research & Development*, 3(1), 228-233.
- [10] Ananth, A., Bhat, S., Naik, R., Suraksha, Pooja Nair, U., & Suraksha. (2021). Parts of speech tagging and extractive summarization techniques for Kannada documents. In *Smart Sensors Measurements and Instrumentation: Select Proceedings of CISCON 2020* (pp. 367-380). Springer Singapore.