

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



Multilingual Text Summarization Using NLP

Mr. Pranav P. Charate, Mr. Shivprasad R. Patil, Mr. Aniket B. Pujari, Mr. Rohan S. Dhere, Mr. Chaitanya B. Pednekar

Department of Computer Science & Engineering KIT's College of Engineering, Kolhapur, India pranav07charate@gmail.com, shivprasadpatil2209@gmail.com, ankitpujari842@gmail.com rohandhere44@gmail.com, pednekar.chaitanya@kitcoek.in

Abstract: In today's information age, the volume of textual data is growing exponentially across various domains, including news, research, and social media. To address the challenge of handling and understanding extensive volumes of text, this project focuses on developing an NLP-based automated text summarization system that efficiently generates concise and accurate summaries. The system supports multiple languages, specifically Hindi, English, and Marathi, ensuring key information is preserved while significantly reducing the overall length and redundancy of the content. To ensure scalability and efficiency, the system is designed to be deployed on cloud infrastructure, leveraging the power of services like AWS or Azure. A simple and intuitive web interface, developed with modern web technologies, enables users to enter text and obtain summaries in their desired language seamlessly

Keywords: Text Summarization, NLP, Pegasus Model, Indic BART, Multilingual

I. INTRODUCTION

In a time when information travels instantaneously, we find ourselves entangled in a vast sea of textual data. From the most recent revolutionary research discoveries and breaking news updates to the myriad thoughts shared on social media platforms, managing and extracting key points from this ocean of information becomes increasingly daunting. Our project rises to this challenge, presenting an intelligent, multilingual NLP-based automated text summarization system.

Our system, carefully crafted to handle Hindi, English, and Marathi, condenses vast amounts of text into clear, concise summaries. By integrating cutting-edge natural language processing techniques and deploying the solution using scalable cloud infrastructure, we deliver fast and precise results through a sleek, user-friendly web interface. This project is not merely about reducing text length; it's about breaking language barriers and enhancing the accessibility of knowledge across diverse linguistic landscapes. Join us on this journey to transform the way we consume information, one summary at a time.

II. LITERATURE REVIEW

Recent Progress on Text Summarization[1]:

The domain of text abstraction has seen so much advancements in recent years, evolving into three primary categories: Extractive, Abstractive, and Hybrid methodologies.

- In extractive summarization, key phrases or sentences are picked directly from the original text. This approach has the advantage of maintaining grammatical correctness and coherence, as it uses the original text without modification. Its simplicity and effectiveness make it widely used, especially in systems requiring quick and reliable summaries.
- On the other hand, abstractive summarization generates new sentences that encapsulate the core ideas of the source text. This method tries to mimic human-level understanding and creativity, making it highly sophisticated. However, it demands more complex processing and advanced NLP techniques to ensure linguistic and contextual accuracy, which remains a significant challenge.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



Hybrid models combine elements of both extractive and abstractive approaches. These models aim to leverage the strengths of both to generate summaries that are both comprehensive and coherent. Recent studies indicate that hybrid models often produce better quality summaries by balancing the precision of extractive methods with the flexibility of abstractive techniques.

Automatic Text Summarization Methods: A Comprehensive Review [2]:

This detailed review paper explores various concepts relevant to automatic text summarization, systematically categorizing summarization methods based on the number of inputs, summary languages, and summary domains.

Summarization Approaches:

- Extractive Summarization: Focuses on selecting specific sentences or phrases from the original text.
- Abstractive Summarization: Generates new sentences that covers the important points of the original text.
- Hybrid Summarization: Combines extractive and abstractive methods.

Techniques Used:

- Statistical Methods: These are based on statistical features of the text, like word occurrence rates and sentence positioning.[9][10]
- Linguistic Methods: Utilize NLP techniques to understand and generate summaries.
- Machine Learning: Machine learning for summarization involves training a system to classify sentences into summary and non-summary sentences using a dataset[8].
- Deep Learning: Advanced neural networks, such as BERT and GPT, are used to create high-quality summaries.

Standard Datasets:

- Several benchmark datasets are widely utilized for developing and evaluating text summarization models. These include:
 - DUC (Document Understanding Conference) datasets: Contain a wide range of news articles used for extractive summarization.
 - CNN/Daily Mail dataset: Popular for training and evaluating abstractive summarization models.

Evaluation Metrics:

- To evaluate the effectiveness of generated summaries, several metrics are utilized, such as:
 - ROUGE (Recall-Oriented Understudy for Gisting Evaluation): Measures the overlap between the generated summary and a reference summary.
 - BLEU (Bilingual Evaluation Understudy): Frequently used in machine translation, also applied to evaluate summarization.

Future Scopes for Research:

- Improved Abstractive Models: Developing models that can better understand and generate human-like language.
- Multilingual Summarization: Creating systems that can summarize text in multiple languages.
- Real-time Summarization: Developing models that can generate summaries on the fly without compromising accuracy.

III. METHODOLOGY

The methodology for our project on text summarization using natural language processing (NLP) used advanced pretrained models that are specially designed for creating summaries. We worked with two models: IndicBART-XLSum by AI4Bharat for summarizing Hindi and Marathi texts, and Pegasus-XSum by Google for English texts. We also created a functionality to convert the summarized text into other languages using Google Cloud Translate v2 API. Below is a step-by-step explanation of our methods.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



3.1 Pretrained Models: IndicBART – XLSum[4]:

Developer: AI4Bharat

- Language Support: IndicBART-XLSum is made for many Indian languages. We focused on using it for Hindi and Marathi.
- Architecture: Indic BART is a type of transformer model. It has two parts: an encoder that reads the input text and a decoder that generates the summary.
- Dataset: The model is trained on the XLSum dataset, which includes high-quality summaries of articles in over 45 languages, including Hindi and Marathi. These articles come from a variety of news topics.
- Training: IndicBART XLSum was trained on a large collection of multilingual data and then fine-tuned for summarizing tasks using XLSum.
- Advantage: This model can understand the unique grammar and style of Hindi and Marathi, producing summaries that are both accurate and natural.

Pegasus - XSum[3]:

Developer: Google Research

- Language Support: Summarizes English texts.
- Architecture: Pegasus uses a transformer framework where it learns to predict entire missing sentences, making it especially good for creating summaries.
- Dataset: The model is fine-tuned on the XSum dataset, which focuses on creating very concise single-sentence summaries of English news articles.
- Training: Pegasus is pretrained on a large dataset of web documents and then fine-tuned on XSum to handle summarization tasks.
- Advantage: Pegasus generates summaries that are clear, meaningful, and highly condensed, while still covering the main points of the input text. According to Google Research [5], PEGASUS employs a novel gap-sentence generation technique that significantly improves the performance of abstractive summarization models.

Data Processing[7]:

- To ensure the input texts were ready for summarization, we performed these preprocessing steps:
- a) Language Identification: We checked the language of each text to categorize it as Hindi, Marathi, or English.
- b) Text Cleaning: Removed unnecessary characters, extra spaces, and metadata from the text.
- c) Tokenization: Broke down the text into smaller pieces (tokens) using tools specific to each model.
- d) Truncation/Padding: Adjusted the length of the input text to meet the requirements of each model.

Summarization Workflow:

- a) Model Loading: Pre-trained models from the Hugging Face Transformers library were utilized.
- b) Inference: The input text was given to the models, which generated summaries.
- c) Translation: Using Google Cloud Translate v2 API, we added the ability to translate the summaries into other target languages if needed.
- d) **Postprocessing:** After the models generated the summaries, we cleaned and adjusted them to ensure they were grammatically correct and easy to read.
- e) Evaluation: We checked the quality of the summaries using both automatic tools and human reviews.



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



Evaluation Metrics:

- a) **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** Evaluates the overlap between the generated summary and the reference summary.
- **b) BLEU** (**Bilingual Evaluation Understudy**): Assesses the fluency and coherence of the generated summary.
- c) Human Evaluation: Experts reviewed the summaries for relevance, clarity, and quality.

Implementation Environment:

- a) Programming Language: Python
- b) Frameworks: Hugging Face Transformers, PyTorch
- c) Hardware: We used NVIDIA GPUs to speed up the process of running and fine-tuning the models.

By using IndicBART-XLSum, Pegasus-XSum, and Google Cloud Translate V2 API, our methodology ensures that we can produce high-quality summaries in Hindi, Marathi, and English, andmake them accessible in other languages. This method enhances the usefulness of summarization for a broader audience.

Frontend Development:

The frontend is developed using React with Vite, a popular JavaScript library for building user interfaces. The project uses Vite as the build tool for a faster and leaner development experience. The following libraries are used to enhance the functionality:

React Router: For client-side routing.

React Toastify: For displaying toast notifications.

Axios: For making HTTP requests to the backend.

Font Awesome: For adding icons to the UI.

Bootstrap: For responsive design and styling.

User Authentication:

User authentication is managed using a custom context. The context provides methods for logging in and logging out, and maintains the authentication state using sessionStorage.

Text Summarization:

The process involves the following steps:

Language Detection: The input text's language is detected using the Google Translate API.

Summarization Request: The detected language, input text, and desired output language are sent to the backend API for summarization.

Display Summary: The summarized text is displayed to the user and can be downloaded as a PDF.

Data Management:

The project uses Axios to interact with the backend API for managing user data and summaries. The following API functions are defined:

User APIs: getUser, addUser, updateUser, deleteUser.

Summary APIs: getSummaries, getSummary, addSummary, updateSummary, deleteSummary.

Styling:

The project uses Bootstrap for responsive design and custom CSS for additional styling. Font Awesome icons are utilized to improve the user interface.



DOI: 10.48175/IJARSCT-25688







Volume 5, Issue 10, April 2025

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal



IV. WORKING



Figure 1. Workflow of Text Summarization and Translation Process

This diagram illustrates the workflow of our project, outlining the sequence of steps involved in text summarization and translation. The process starts with the input text, which undergoes several stages including processing, feature extraction, and summarization. The generated summary is subsequently translated into the desired language, producing the final output. Every step plays a vital role in maintaining the accuracy and quality of both the summaries and translations.



Figure 2. Flowchart

1) IndicBART – XLSUM Architecture:

IndicBART-XLSum is a transformer-based model designed specifically for summarizing texts in various Indian languages. The architecture is made up of two primary components:

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688







International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



a) Encoder:

• The encoder analyzes and interprets the input text. It transforms the text into a sequence of hidden states that represent the input's meaning and context.

b) Decoder:

- The decoder generates the summarized text based on the hidden states produced by the encoder.
- It uses attention mechanisms to focus on relevant parts of the input text while generating the summary.

2) Pegasus – Xum Architecture:

Pegasus-XSum is also a transformer-based model, optimized for summarizing English texts. Its architecture includes:

a) Encoder:

- The encoder interprets and handles the input text.
- It produces hidden representations that encapsulate the semantic meaning and context of the input.

b) Decoder:

- The decoder generates summaries by predicting entire missing sentences within the input text.
- It employs attention mechanisms to focus on the most important parts of the text for summarization.

V. EVALUATION

✓ Performance of IndicBART-XLSum[6]: IndicBART-XLSum is a highly efficient model specifically designed to handle Indic languages. Its performance on different tasks can be measured based on speed, accuracy, and efficiency.

Speed: IndicBART-XLSum processes text at an impressive rate. Compared to other models such as mBART and mT5, IndicBART-XLSum is notably smaller in size. This compactness results in lower computational requirements for fine-tuning and decoding. Consequently, the model can execute tasks rapidly without compromising on performance.

Accuracy: IndicBART-XLSum demonstrates remarkable accuracy across several Indic languages, including Bengali, Gujarati, Hindi, Marathi, Punjabi, Tamil, and Telugu. The model's performance is assessed using the Rouge-1, Rouge-2, and Rouge-L scores, which gauge its ability to generate coherent and contextually relevant text. The table below presents a detailed analysis of these scores for the Hindi and Marathi languages:

	PERFORMAN	CE OF (LSUM	
LANGUAGE	ROUGH - 1	ROUGH - 2	ROUGH - L
Hindi (hi)	0.220394	0.065464	0.198816
Marathi (mr)	0.172568	0.062591	0.160403

Figure 3. INDICBART-XLSUM Evaluation Matrix

The table presents the Rouge-1, Rouge-2, and Rouge-L scores achieved by IndicBART-XLSum for Hindi and Marathi languages. These scores are critical in evaluating the model's performance[11]:

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



Rouge-1: Evaluates the match of individual words (unigrams) between the generated text and the reference text.

Rouge-2: Evaluates the match of word pairs (bigrams) between the generated text and the reference text. **Rouge-L:** Measures the longest common subsequence between the generated text and reference text, emphasizing the model's ability to maintain the sequence of ideas.

The scores highlight IndicBART-XLSum's capability to generate text that closely aligns with human-written content, showcasing its potential in various applications requiring natural language understanding and generation in Indic languages.

✓ Performance Pegasus – XSum

The Pegasus Model exhibits impressive performance across various datasets, effectively summarizing texts while preserving key information. Let's explore its performance, speed, accuracy, and efficiency in detail.

Speed: The model's training time is noteworthy, clocking in at a total of 1.5 million steps, which is three times the original 500,000 steps. This extended training period allows the model to converge more slowly and accurately, resulting in enhanced performance.

Accuracy: The Pegasus Model achieves high accuracy in summarization tasks, with ROUGE scores (a measure of summarization quality) ranging from 43.90 to 59.67 across different datasets. These scores indicate the model's ability to condense long pieces of text into concise and informative summaries. The table below highlights the ROUGE scores achieved by the model on various datasets:

	PERFORMAN PEGASUS – 2	CE OF XSUM	
DATASET	ROUGH - 1	ROUGH - 2	ROUGH - L
C4	45.20	22.06	36.99
HugeNews	47.21	24.56	39.25
		10.70	24.70

Figure 4. PEGASUS-XSUM Evaluation Matrix

The table provides the ROUGE-1, ROUGE-2, and ROUGE-L scores for the Pegasus Model across three different datasets: C4, HugeNews, and WikiHow. These scores are critical in evaluating the model's summarization performance[12]:

ROUGE-1: Assesses the overlap of single words (unigrams) between the generated summary and the reference summary.

ROUGE-2: Assesses the overlap of consecutive word pairs (bigrams) between the generated summary and the reference summary.

ROUGE-L: Measures the longest common subsequence between the generated summary and reference summary, emphasizing the model's ability to maintain the sequence of ideas.

The scores underscore the model's capability to generate summaries that closely align with human-written content, showcasing its potential in various applications requiring precise and coherent text summarization.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



V. OUTPUT

	🔄 sporsavenstan 🔮 kapanta 🔮 etti avetajenisto. 🔿 stalik kaji 📳 endinetnalima 🔍 teorrajintorian. 📳 ninstrukvalta.	
GenZ Intelligents		Fana Lagia Bagir
	Welcome to Our Multilingual Text Summarization System!	
	In an ris where information is both abuildord and fregmentatic our automated test auronations operant offers a vital adultor. Designed to condense lengthy tests into concess informative auronatios, care system supports multipla lenguages including heids. (Explain our Montell Western Science of the content queck) and efficiently. Learning a timing align entantic language incomessing includings, or plation anise to rease track elemention accessible to services, tracked or gluing align barriers and enhancing understanding. Explore here our lystem on function you demote consistence experience.	
	9 2021 Con?EtraHarsta Al repta received.	

Figure 5. Landing Page Screenshot

The Landing Page serves as the entry point to our text summarization and translation application. It provides a brief overview of the service, highlighting its key features and benefits. From this page, users can access other sections of the application.

- @ Suchasayan X a				
(NAMA) EB — Indephasing to 🖄 Contained Here 💐 Majarta 🧯	Web-Development Cou-	transform E George better 10 feet	repodorom. 💼 nevezorateren.	
		Register		
	Full Name	enter your fuil name		
	Email	Trite gaarenal		
	Phone	Criter your phone number		
	Number			
	Gender	Select Greder		
	Address	True your address		
	Pessword	Writer your pacoword	200	
	Confirm	Confirm your personnel	200	
	Password			
		Submit		
		Missedy have an account? Login		

Figure 6. Registration Page Screenshot

The Registration Page enables new users to set up an account. It captures essential information such as username, email, and password. The page ensures a smooth onboarding process with a simple and intuitive design.



Figure 7. Login Page Screenshot

The Login Page allows registered users to sign in to their accounts by entering their email and password. It also provides options for password recovery and includes a link to the registration page for new users.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal



Volume 5, Issue 10, April 2025

х фоксание х						
			🛛 Cruistchan 📲 Cecars Debtord	and the state of t		🗆 x Geotrate
GenZ Intelligents					🕑 Logis successful	×
			Dashboard			
				-		
	Inp	it Language	Select Language			
	Out	put Language	Select Language			
	Inp	at Text				
	in	tor your test				
			Scenario			

Figure 8. Dashboard Page Screenshot

The Dashboard Page provides users with an overview of their activity and access to various functionalities. Users can view summaries they have generated, track their history, and access settings. The page offers quick access to essential features, enhancing user experience.

× фактана х н		- a x
e -> @ Q InshortST/Unitery		
(Remain) 22 - The Referencements in	n, 🖷 Skanta 😫 Wellewissen (k) Castellien 📱 Gezen bettert 🕫 Sectorabation. 🏥 neutzestermen.	
GenZ Intelligents		Home Dashboard History Logant
	History	
Date: 1/2/	2024	
	The global shift towards ensembles merge courses is incuring our own it is such to reduce callose existions and context climate charge tencovations is strate, wind, and hydramentic powers in making provided a single many efficient and cast effective. Governments and	
	reproductions are investing Rearily in accorrection logistic oceans autoimable one are solutions for the house.	
	© 2024 Cen2 Medigence. All highes reversed. f f f © in	

Figure 9. History Page Screenshot

The History Page displays a list of all the text summaries that the user has generated. Each entry includes the original text, the summarized output, and the date it was created. Users can easily revisit and review their past summaries.







International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



GenZ Intelligents	
Previo	bus Summary
Date : 11/2/2024	
Input Language	English ~
Output Language	English ×
Input Text	
The global shift tov accelerating as cou emissions and com solar, wind, and hy renewable energy i Governments and i green technologies solutions for the fu	vards renewable energy sources is ntries seek to reduce carbon bat climate change. Innovations in droelectric power are making more efficient and cost-effective. organizations are investing heavily in . to create sustainable energy ture.
	Save Changes
Summarized Text	
The global transitio momentum, driven emissions and tack wind, and hydroele and cost-effectiven green technologies	in to renewable energy is gaining by the need to reduce carbon le climate change. Advances in solar, ctric power are enhancing efficiency ess, with significant investments in s.
Download Sun	mary Remove Summary
© 2024 Geni	Z Intelligents. All rights reserved. f ¥ @ in

Figure 10. Specific Summary Details Page Screenshot

The Specific Summary Details Page provides detailed information about a particular summary. It includes the input text, the generated summary, and any translated versions. Users have the functionality to update the input text and generate a new summary for it. Additionally, this page allows users to delete summaries and download them in PDF format as shown in Figure 10 This page offers comprehensive management and customization options for each summary.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



A B REINER X B RESERVOIRESEE	NC X F	
	🗄 blaseta 🔮 Vel-Sedamento, 👋 Scottinan 🚦 Gesen Settard 🕫 Sentrastran, 📫 mestamberren.	
Entering the international and the second	1 F1 - 1046 + 🗓 🚸	
	Summary Details	
	Date: 11/2/2024	
	User Email: sp1788771@gmail.com	
	Input Language: English	
	Output Language: English	
	Input Text:	
	The global shift broads removable energy sources is accelerating as counties sould be reduce earbon entraison and contrait division thango. Innovations as using using and bydroekolic prover are making removable and invasing heaving in green technologies to create sustainable energy solutions for the future.	
	Output Text:	
	The global transition to reveable energy is gaining momentum, driven by the need to neutral earthon ensistence and tackle driven change. Advances in solar, which and hydrolearth power are enhancing efficiency and cost-effectiveness, with significant investments in green technologies.	

Figure 11. Download PDF format Screenshot.

VI. CONCLUSION

The multilingual text summarization system developed in this project successfully addresses the challenge of processing and condensing large volumes of text while preserving key information. By leveraging advanced NLP models like IndicBART-XLSum for Hindi and Marathi and Pegasus-XSum for English, the system generates highquality summaries with enhanced readability and accuracy. The integration of Google Cloud Translate v2 API further extends the usability of the system by allowing translations into multiple languages. Through rigorous evaluation using ROUGE and BLEU metrics, the project has demonstrated the effectiveness of its summarization approach, while the user-friendly React-based web interface enhances accessibility. This project not only improves information accessibility across languages but also serves as a foundation for future advancements in real-time summarization, enhanced abstractive techniques, and multilingual NLP applications, contributing to bridging linguistic gaps and optimizing the way users interact with textual data in the digital era.

REFERENCES

- [1] Suad Alhojely and Jugal Kalita, "Recent Progress on Text Summarization," 2020 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1503-1509, 2020, doi: 10.1109/CSCI51800.2020.00278.
- [2] Divakar Yadav, Jalpa Desai, and Arun Kumar Yadav, "Automatic Text Summarization Methods: A Comprehensive Review," International Journal of Computer Applications, 2021.
- [3] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu, "PEGASUS: Pre-training with Extracted Gapsentences for Abstractive Summarization," arXiv preprint arXiv:1912.08777, 2020. DOI 10.48550/arXiv.1912.08777.
- [4] Raj Dabre, Himani Shrotriya, Anoop Kunchukuttan, Ratish Puduppully, Mitesh M. Khapra, and Pratyush Kumar, "IndicBART: A Pre-trained Model for Indic Natural Language Generation," arXiv preprint arXiv:2109.02903, 2022. DOI: 10.48550/arXiv.2109.02903.
- [5] Peter J. Liu and Yao Zhao, PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization
- [6] Arjit Agarwal, Soham Naik, and Sheetal Sonawane, "Abstractive Text Summarization for Hindi Language using IndicBART," Forum for Information Retrieval Evaluation (FIRE), CEUR Workshop Proceedings, vol. 2022, pp. 1-7, 2022.
- [7] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, O'Reilly Media, 2009. (Covers text cleaning, tokenization, and preprocessing steps in NLP.)
- [8] R. Oak, "Extractive techniques for automatic document summarization: a survey," International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, no. 3, pp. 4158–4164, 2016
- [9] M. Gambhir and V. Gupta, "Recent automatic text summarizationtechniques: a survey," Artificial Intelligence Review, vol. 47, no. 1, pp.1-66, 2017.

Copyright to IJARSCT www.ijarsct.co.in



DOI: 10.48175/IJARSCT-25688





International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 5, Issue 10, April 2025



- [10] V. Gupta and G. S. Lehal, "A survey of text summarization extractivetechniques," Journal of Emerging Technologies in Web Intelligence, vol. 2, no. 3, pp. 258–268, 2010.
- [11] Raj Dabre, Himani Shrotriya, Anoop Kunchukuttan, Ratish Puduppully, Mitesh M. Khapra, Pratyush Kumar, IndicBART: A Pre-trained Model for Indic Natural Language Generatio
- [12] Pegasus-Xsum, Google, https://dataloop.ai/library/model/google_pegasus-xsum/
- [13] IndicBART, AI4Bharat, https://dataloop.ai/library/model/ai4bharat_indicbart/

Copyright to IJARSCT www.ijarsct.co.in



