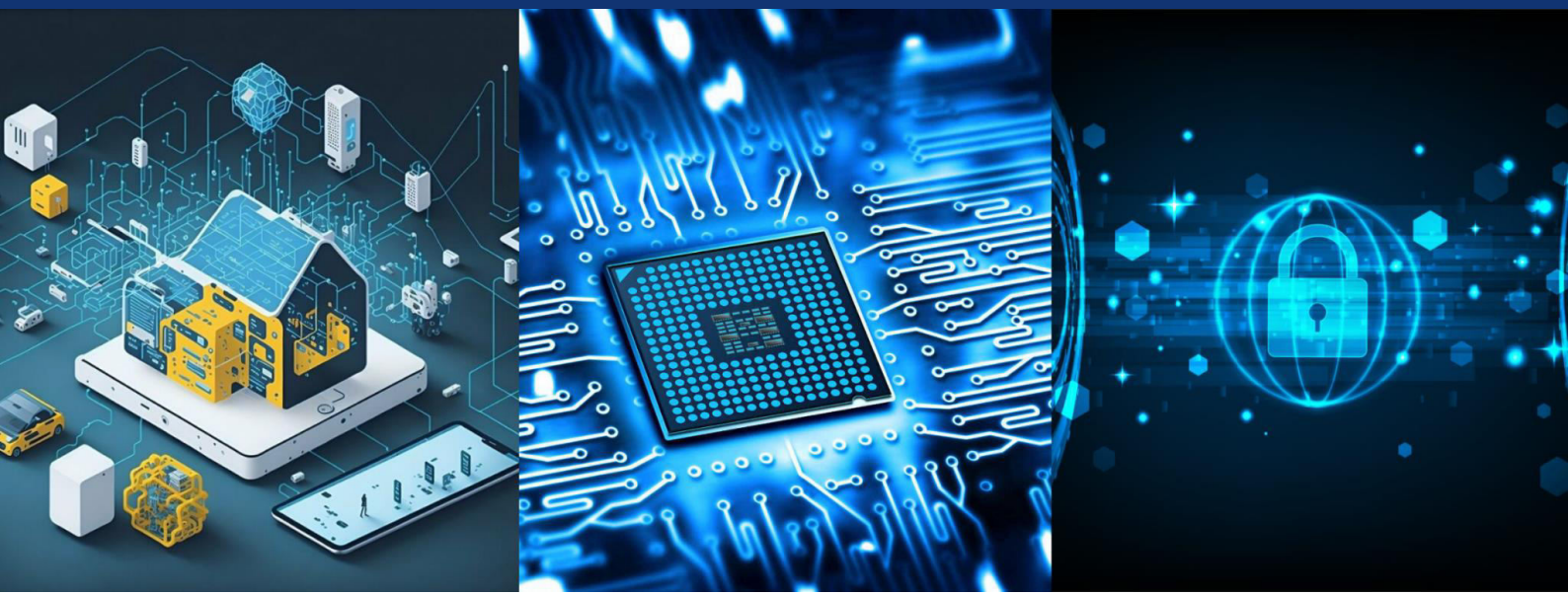
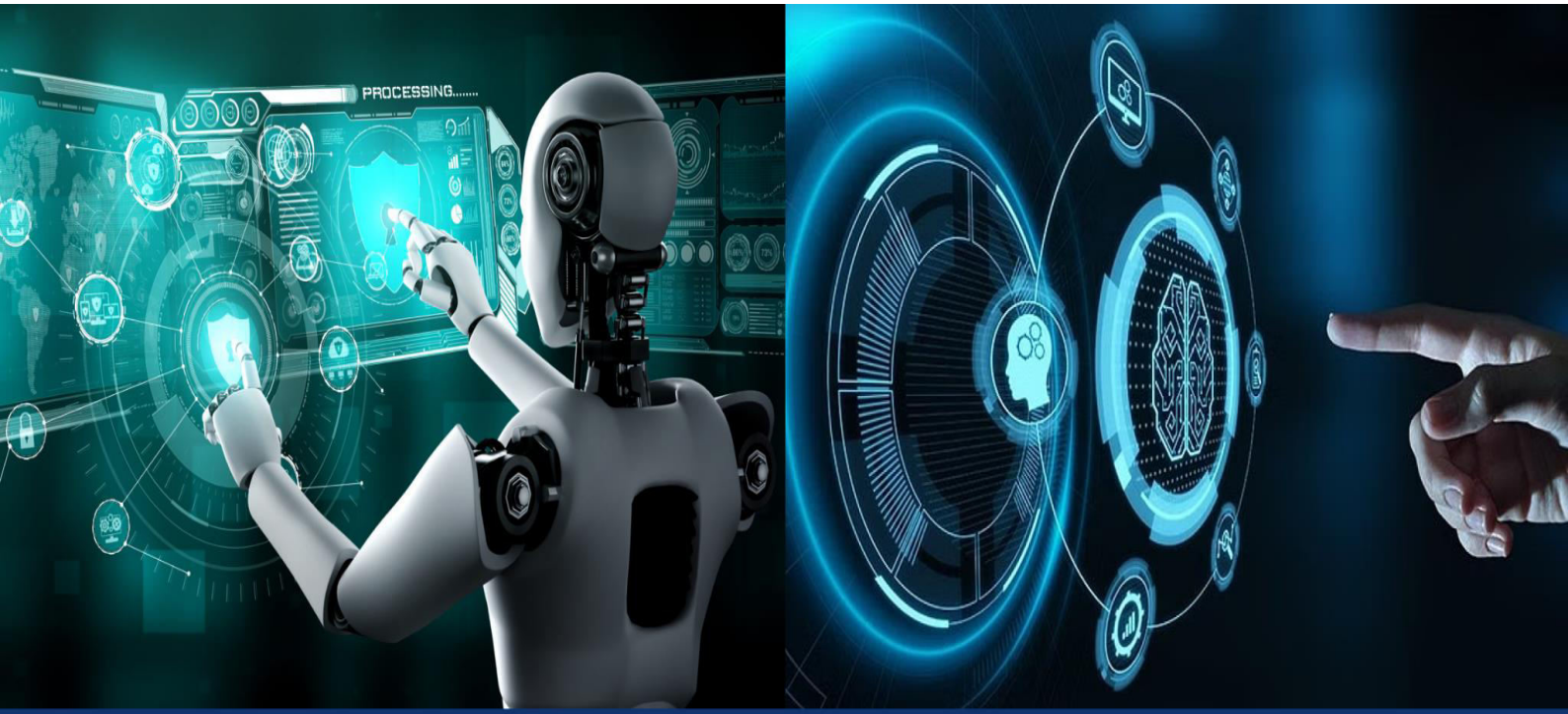
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 13, Issue 5, May 2025**

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305251

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Text Summarization: Comparing the Summarization styles of different model

Nitesh Rajpurohit[1], Prof. Ashwini Garkhedkar[2]

Student, P.E. S. Modern College of Engineering, Pune, Maharashtra, India[1]

Professor, P.E. S. Modern College of Engineering, Pune, Maharashtra, India[2]

ABSTRACT: Summarization of text is an important aspect of natural language processing (NLP), aimed at generating short, consistent summaries of long documents without sacrificing important information. Transformer models have transformed automatic summarization. This paper compares two top-performing CPU-friendly models, T5-Base and DistilBART, on three genres—Entertainment Finance, and Technology. The research paper compares the performance of the two models on ROUGE score(quantitative assessment) and qualitative assessment based on the effectiveness and efficiency of the models in summarizing content across genres. The results reveal varying degrees of summarization accuracy based on genre, highlighting the capability and limitation of the models.

KEYWORDS: Transformers, T5-Base ,DistilBART , Text Summarization

I. INTRODUCTION

The exploding growth of content online has made automated text summarization a must-have method for efficient information gathering. Extractive methods that take out key sentences from the body of text do not often carry the original essence and rationale behind the content. On the contrary, abstractive summarization, powered by transformer-based technology, generates summary versions by rephrasing the input in more natural and reduced form.Transformer models have transformed NLP applications, including summarization. BERT, GPT, and T5 are some of the models that have proved to have incredible text understanding and generation capabilities. Here, we discuss two Transformer-based models, T5-Base and DistilBART, which strike a balance between performance and computational load. They are particularly fine-tuned for summarization but are lightweight and CPU-efficient compared to large models like Pegasus or GPT-4.This research examines T5-Base and DistilBART summarization performance across three content genres—Entertainment, Finance, and Technology—to verify how well they can process different types of content.

II.LITERATURE SURVEY

Text summarization is a high-profile problem in the field of Natural Language Processing (NLP), with research activity dating back to the early rule-based and extractive approaches. Previous approaches, such as LexRank and TextRank, employed graph-based ranking algorithms to rank significant sentences in a document. While these approaches were suitable for extractive summarization, they were unable to generate readable, human-written summaries.

The advent of deep learning revolutionized summarization techniques with the introduction of sequence-to-sequence (seq2seq) models, which allowed for abstractive summarization. Pointer-Generator Networks were among the advancements that improved coherence by employing copying mechanisms, which enabled handling out-of-vocabulary words[3]. The models, however, still struggled with factual errors.

With the introduction of transformer-based models, BERTSUM, T5, and BART have turned out to be state-of-the-art for summarization[1]. T5 (Text-to-Text Transfer Transformer) converts every NLP task into text-to-text, thereby enabling flexible summarization[4]. DistilBART, the distilled version of BART, achieves a balance between performance and efficiency by reducing computational complexity without compromising robust summarization capabilities[2].Despite these advancements, model performance varies based on the domain and content type This research aims to fill this gap by comparing T5-Base and DistilBART on various genres—finance, entertainment, and technology—to determine their performance in creating high-quality summaries.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305251

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III.METHODOLOGY

This paper explores the summarization capability of T5-Base and DistilBART on three genres of articles: Entertainment, Finance, and Technology. The corpus is made up of articles borrowed from a diverse selection of sources, thus ensuring adequate representations of topics. Both models summarize each article, and the output acquired is measured in terms of quantitative and qualitative metrics.

3.1. Models Employed:
T5-Base: A transformer model that transforms text tasks into a text-to-text format that is trained from large-scale summarization data. It performs reasonably well in generating coherent and informative summaries with computational efficiency[4].DistilBART is a variant of the BART model, which is reduced in size without losing its ability to generate quality summaries. DistilBART is designed to be more efficient and have faster inference times[2].

3.2. Assessment Criteria: The summarization model quality is measured by executing ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores[5]:
ROUGE-1: Calculates the overlap of unigrams (single words) between the reference summary and the automatically produced summary.
ROUGE-2: Scores the overlap of bigrams (two adjacent words), more contextually similar.
ROUGE-L: Computes the longest common subsequence, reflecting structural similarity between summaries.

By comparing the summarization outputs on a broad spectrum of genres, this paper tries to identify the strengths and weaknesses inherent in each model in processing various types of inputs. Furthermore, we discuss the pragmatic advantages of employing light models such as T5-Base and DistilBART in practical applications where the use of computational resources is a primary concern.

IV.RESULT AND DISCUSSION

4.1. ROUGE Score Analysis

To evaluate the performance of the T5-Base and DistilBART models in summarizing text, we computed ROUGE scores for three categories: Entertainment, Finance, and Tech.
The results of average ROGUE scores across genres are given in the following table.

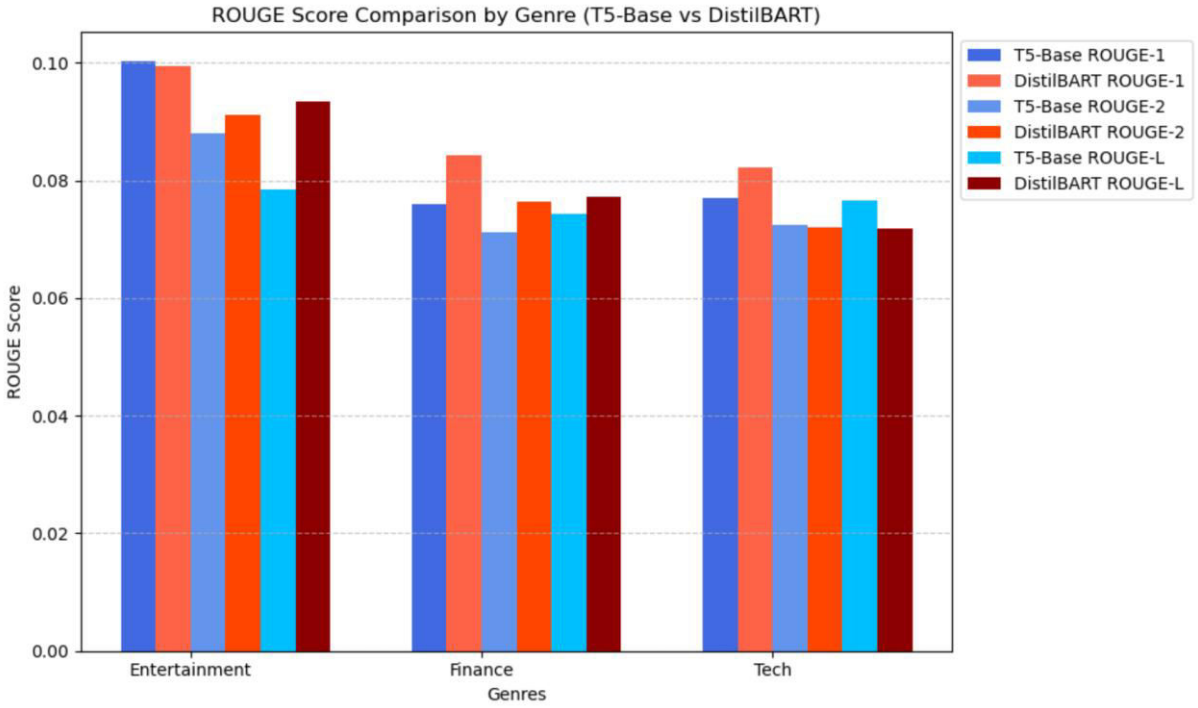
Genre	Model	ROUGE-1	ROUGE-2	ROUGE-L
Entertainment	**T5-Base**	0.1002	0.0881	0.0785
	DistilBART	0.0995	0.0912	0.0934
Finance	**T5-Base**	0.0759	0.0709	0.0743
	DistilBART	0.0843	0.0765	0.0772
Tech	**T5-Base**	0.0770	0.0724	0.0766
	DistilBART	0.0823	0.0711	0.0718

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305251

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Entertainment Articles:

- DistilBART achieved slightly higher ROUGE-2 and ROUGE-L scores when compared to T5-Base, which suggest that it captures b phrase-level and sentence-level coherence better for entertainment-related content.
- T5-Base was not far behind, which showed its competitive performance in summarizing conversational and narrative-driven content.

Finance Articles:

- DistilBART was far ahead of T5-Base in all ROUGE metrics, which suggest that it was more effective at summarizing structured financial data.
- The complex and numerical nature of financial content have played a role in the slight performance gap between two models

Tech Articles :

- The two models were close to each other, with DistilBART leading in ROUGE-1 but slightly behind in ROUGE-2 and ROUGE-L.

- This suggests that both models perform similarly on technical content, but neither with a clear advantage.

4.2. Qualitative Observations

- **Summary Length**: DistilBART produced longer summaries than T5-Base, which explains its higher ROUGE scores in many cases.
- **Content Retention**: DistilBART sometimes included additional words and context, making its summaries more detailed. However, this led to occasional redundancy.
- **Fluency vs. Conciseness**: T5-Base summaries were more concise in nature but sometimes did not include crucial details. DistilBART was more descriptive but occasionally included unnecessary information in the summaries.

V. CONCLUSION

This research assessed the summarization ability of the T5-Base and DistilBART models on three different genres: Entertainment, Finance, and Technology. . Based on ROUGE scores and qualitative analysis, findings from research are:

- DistilBART surpassed T5-Base in Entertainment and Finance and generated more elaborate summaries.
- T5-Base was comparable in the Tech category, with no clear winner for technical content.
- DistilBART summaries were generally longer and also more context-complete, although T5-Base produced relatively shorter outputs.
- DistilBART faced the issue of hallucination on one of the article and produced completely irrelevant content.
- The results show that DistilBART may be more appropriate for summarizing in-depth or factual material, while T5-Base appears to be more efficient in conciseness. Both models are capable enough considering their cup friendliness .

Future enhancement would include expanding the dataset to a larger collection of articles and experimenting with different summary length constraints.

REFERENCES

1. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
2. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv:1910.01108*, 2019.
3. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017
4. Kotte, K. R., & Panyaram, S. (2025). Supply Chain 4.0: Advancing Sustainable Business. Driving Business Success Through Eco-Friendly Strategies, 303.
5. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1-67, 2020.
6. C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," *Proc. Workshop on Text Summarization Branches Out (WAS 2004)*, Barcelona, Spain, Jul. 2004

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details