iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** I **Month of publication:** January 2024

DOI: https://doi.org/10.22214/ijraset.2024.56564

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Text Summarization Using NLP

Mr. Harshal Bharati[1], Mr. Premchand Tarange[2], Mr. Ankit Kumar[3], Mr. Shubham Mate[4], Prof. R.C. Bhagananagre[5]
[1, 2, 3, 4]*Student dept. of Computer Engineering SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra*
[5]*Asst. Professor, dept. of Computer Engineering, SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra*

Abstract: ***This Project represents the work associated with Text Summarization. In this paper, we present a framework for summarizing the massive facts. The proposed framework depends on highlight extraction from internet, using each morphological element and semantic data. Presently, in which large facts is available on the internet, it's far maximum important to offer the improved approaches to extract the statistics quickly and most successfully. It may be very hard for human beings to manually extract the precis of a large document of text. There are lots of text substances to be had on the internet. So, there's a hassle of looking for related files from the quantity of documents to be had and absorbing associated statistics from it. In essence to determine out the previous issues, the automated textual content summarization could be very a whole lot essential. Text Summarization is the method of figuring out the most vital and significant information in a input report or set of related input files and compressing all the inputs into a shorter version preserving its normal goals.***
Keywords: Machine Learning, Text Summarization, Natural Language Processing (NLP), Tokens

I. INTRODUCTION

In this paper, we present a framework for Text Summarization. The proposed framework depends on summarizing the text from net, utilizing both morphological factors and semantic records. The period of textual content statistics is growing, and people have less time to examine that information. The Internet, media and other facts assets have a large sell off of records and therefore a gadget is required for producing simpler and brief form of information. So, a tool is required for the customers, which would ease out the attempt for them and to study the entire text or remember. Such a systems or tools would be useful and a exceptional time saver for the users. Hectic schedules made it not possible for all of us to study and get admission to the statistics from News information, biographical statistics or from other journals. A reliable and simpler information is wanted to be green. With summaries, People can make efficient choices right away. The motivation right here is to build such a tool which is efficient and creates summaries automatically. Natural Language Processing (NLP) is a area of computerized cogitation in which PCs probe, recognize, and get significance from human language in a radiant and beneficial manner. In addition to typical word processor jobs that deal with messages, such as basic image placement, natural language processing (NLP) considers the several ways that language is constructed: a few words constitute a declaration, a few declarations constitute a sentence, and sentences reveal ideas. NLP expert John Reeling of the software solutions business "Meltwater Group" explained in his article, "How Natural Language Processing helps Uncover Social Media Sentiment." By breaking down the language Because of its significance, NLP frameworks play a lengthy and comprehensive beneficial representation, controlling punctuation, changing shifting the focus of the talk to the message and translating between dialects as necessary. NLP is used to break down the text so that computers can understand human communication. Fair applications such as programmed message outlines, judgement investigations, topic subdivision, named element acceptance, grammatical feature classification, relationship production, stemming, and the list goes on thanks to this human-PC partnership.

II. HISTORY

The practice of reducing a huge volume of text while maintaining its vital information and general meaning is known as text summarizing. For decades, scholars and practitioners have been interested in this topic because of the necessity to efficiently process large volumes of textual data. To meet the issues of text summarization, numerous methodologies and strategies have emerged throughout time.

Early Methodologies (1950s-1990s): Text summarizing attempts date back to the 1950s, when scholars investigated rule-based systems for extracting essential lines from manuscripts. Hans Peter Luhn pioneered the notion of keyword extraction in the 1960s, where keywords are found and utilized to generate document summaries. With early work on text reduction, sentence extraction, and sentence weighting, automated summarization gained traction in the 1970s.

©IJRASET: All Rights are Reserved

Extractive Summarization (from the 1990s to the 2000s): In the 1990s, the extractive summarizing technique, which includes choosing relevant lines or paragraphs from the original material, gained popularity.

To identify sentence relevance, researchers began experimenting with statistical and machine learning techniques, such as word frequency and graph algorithms. Through common objectives and benchmark datasets, the Document Understanding Conference (DUC) and the Text Analysis Conference (TAC) performed critical roles in promoting extractive summarization. Abstractive Summarization (from the 2000s to the present): Abstractive summarizing, in which the summary is formed by rephrasing and synthesizing text, gained popularity in the 2000s as NLP methods advanced. Techniques such as phrase compression and syntactic parsing were employed in early abstractive methods.

In the 2010s, the introduction of neural networks and deep learning revolutionized abstractive summarization, allowing for more fluent and human-like summaries. Pre-trained language models like BERT and GPT-2 have enhanced the quality of abstractive summaries even further.

III. LITERATURE SURVEY

Automatic text summarization approaches (A. T. Al-Taani). In this paper ATS (automated text summarization and the approaches of single document and multi- documents text summarizations have been discussed based on requirements extractive sumarization is used.[1]

Automatic text summarizer (A. P. Patil, S. Dalmia, S. Abu Ayub Ansari, T. Aul and V. Bhatnagar). The aim in this paper was to design and construct an algorithm that can summarize a document by extracting key text and modifying this extraction using a thesaurus. Mainly to reduce the size, maintain coherence.[2]

Text Summarization: A Review (S. Biswas, R. Rautray, R. Dash and R. Dash). In this paper the objective is to explore text summarization by using various technologies and methodologies in creating a coherent summary which including the key points of the original input document.[3]

An overview of Text Summarization techniques (N. Andhale and L. A. Bewoor). This paper gives an overview survey on both extractive and abstractive approaches.[4]

Text Summarization: An Essential Study (P. Janjanam and C. P. Reddy). This paper focuses on the study of abstractive text summarization approaches and the state of art machine learning models used to summarize single and multi- documents and eventually leading to large document summarization.[5]

Natural Language Processing (NLP) based Text Summarization (I. Awasthi, K. Gupta, Bhupendra Jogi, S. S. Anand and P. K. Soni).In this paper study of extractive and abstractive text summarization method is done .It uses linguistic and statistical characteristics to calculate the implications of sentences. This paper also aims at less repetition and accurate summary.[6]

IV. METHODOLOGY

1) *Step in Pre-processing:* One procedure that comes before translation is pre-processing. The summarizer system's input is a document or group of linked documents. The document has to be moved into a bag of words or phrases. Natural Language Processing (NLP) stages like phrase segmentation, tokenization, stop word removal, and stemming are included in the pre-processing step. Following the completion of pre-processing, each token's word frequency and reverse document frequency values are determined.
2) *Sentence Segmentation:* The technique of breaking apart a string of written language into its unit or module sentences is known as sentence segmentation. Languages like English and a few others use punctuation, and using symbols like full stops and period characters are a logical estimation.
3) *Tokenization:* Tokenization is the technique of breaking down phrases into separate tokens that are adjusted by the spaces and may be utilized for further understanding and improvement. Discrete words, phrases, keywords, IDs, etc. can all be used as tokens. Tokens or words are separated throughout the tokenization process by line breaks, punctuation, and white space. Depending on the demands, the punctuation or white space may or may not get intertwined.
4) *Elimination of Stop Words:* Stop Words are terms that are commonly used in the language. Eliminating stop words involves getting rid of terms like "the," "to," "are," "is," and so on. Additionally, stop words are eliminated to improve phrase search support.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

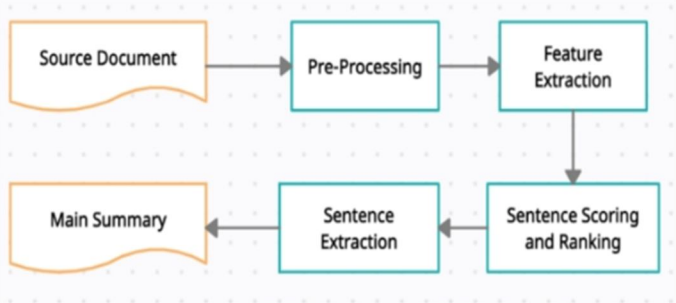


Fig: Step wise approach of text summarization

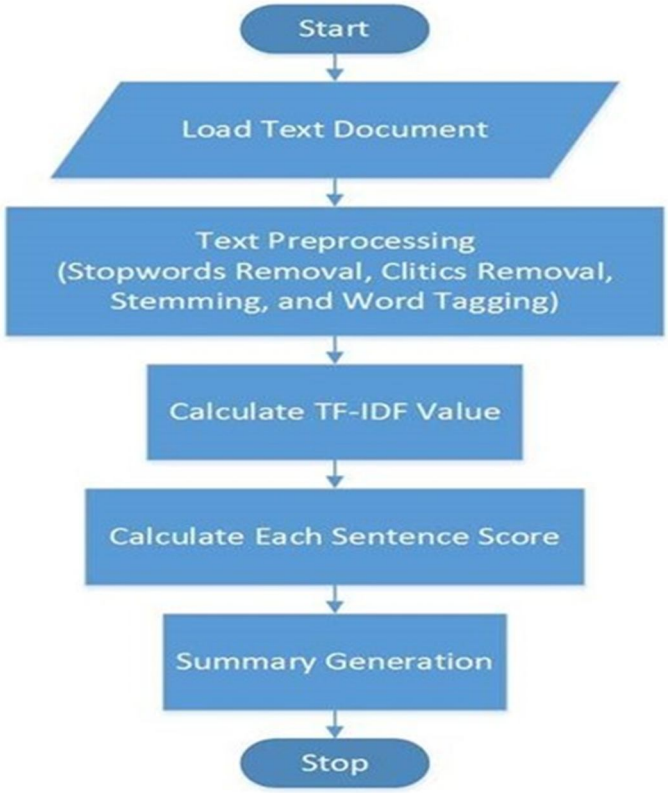


Fig : Work Plan

V. OVERVIEW OF TEXT SUMMARIZATION

Text summarization is the process of condensing a large and often complex piece of text, such as an article, document, or news story, into a shorter version while retaining its essential information and meaning. It tries to give readers with a condensed version of the original text, allowing them to comprehend the essential ideas and key points without having to read the full document.

A. Key Features of Text Summarization

1) *Content Reduction:* Text Summarization decreases the amount of text by removing unnecessary details, repetitious information, and less significant stuff. This function is especially useful for saving time and improving content consumption.
2) *Preservation of Meaning:* Effective text summarizing guarantees that the summary keeps the original text's key meaning and context. This is accomplished by selecting the most significant concepts, ideas, and facts to mention in the summary.
3) *Machine Learning and Natural Language Processing (NLP):* To increase the quality of summaries, several recent text summarizing approaches make use of machine learning and Natural Language Processing (NLP) technology. Abstractive summarization has evolved greatly thanks to pre-trained language models such as BERT and GPT-3.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VI. GUI OF TEXT SUMMARIZATION WEB APPLICATION

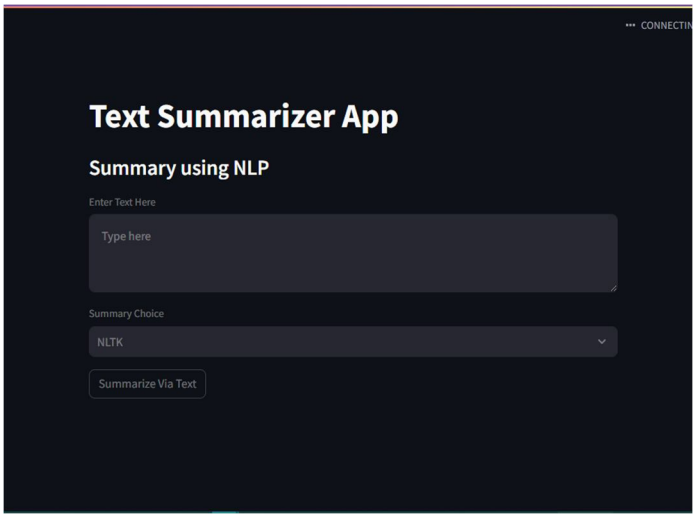


Fig : GUI of Application

VII. CONCLUSION

Text summarization is a challenging task in natural language processing (NLP), but it is also a very useful one. It can be used to shorten long documents, making them easier to read and understand. It can also be used to extract key information from documents, which can be helpful for tasks such as question answering and information retrieval. In this project, we have developed a text summarizer using NLP techniques. Our summarizer is based on the extractive approach, which extracts important sentences from the original text to form the summary. We used a variety of NLP techniques to identify important sentences. We evaluated our summarizer on a standard text summarization dataset, and it achieved good results. Our summarizer was able to generate summaries that were concise and informative, while still preserving the key information from the original text

REFERENCES

[1] T. Al-Taani, "Automatic text summarization approaches," 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 2017, pp. 93-94, doi: 10.1109/ICTUS.2017.8285983.
[2] A. P. Patil, S. Dalmia, S. Abu Ayub Ansari, T. Aul and V. Bhatnagar, "Automatic text summarizer," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 1530-1534, doi:10.1109/ICACCI.2014.6968629.
[3] S. Biswas, R. Rautray, R. Dash and R. Dash, "Text Summarization: A Review," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), 2018, pp. 231-235, doi: 10.1109/ICDSBA.2018.00048. [4] N. Andhale and L. A. Bewoor, "An overview of Text Summarization techniques," 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 2016, pp. 1-7, doi: 10.1109/ICCUBEA.2016.7860024.
[4] P. Janjanam and C. P. Reddy, "Text Summarization: An Essential Study," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi:10.1109/ICCIDS.2019.8862030.
[5] I. Awasthi, K. Gupta, Bhupendra Jogi, S. S. Anand and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1310-1317, doi:10.1109/ICICT50816.2021.9358703.
[6] H. T. Le and T. M. Le, "An approach to abstractive text summarization," 2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR), 2013, pp. 371-376, doi:10.1109/SOCPAR.2013.7054161.
[7] P. R. Dedhia, H. P. Pachgade, A. P. Malani, N. Raul and M. Naik, "Study on Abstractive Text Summarization Techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-8, doi:10.1109/ic-ETITE47903.2020.087.

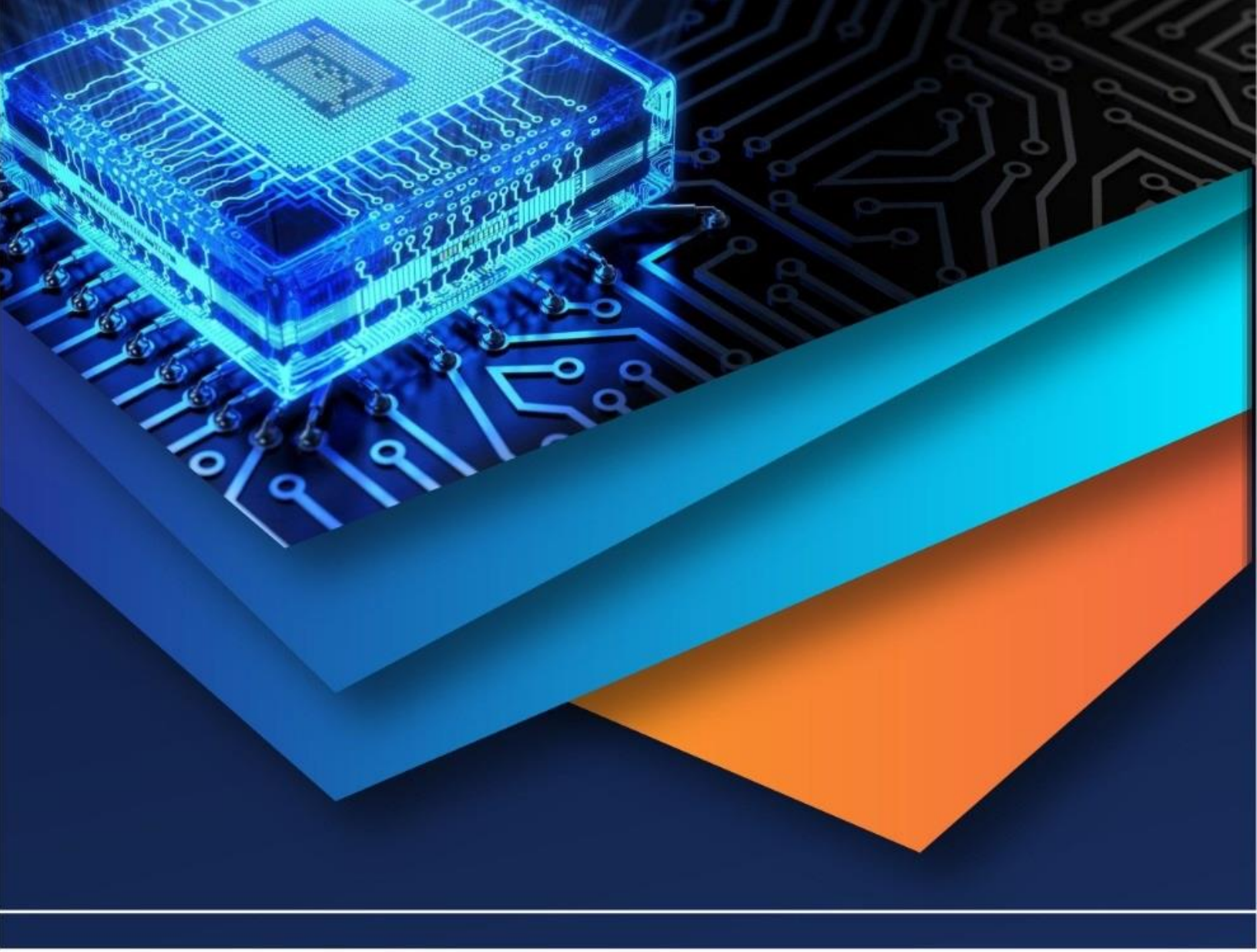
REFERENCES

[1] T. Al-Taani, "Automatic text summarization approaches," 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 2017, pp. 93-94, doi: 10.1109/ICTUS.2017.8285983.
[2] A. P. Patil, S. Dalmia, S. Abu Ayub Ansari, T. Aul and V. Bhatnagar, "Automatic text summarizer," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 1530-1534, doi:10.1109/ICACCI.2014.6968629.
[3] S. Biswas, R. Rautray, R. Dash and R. Dash, "Text Summarization: A Review," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), 2018, pp. 231-235, doi: 10.1109/ICDSBA.2018.00048.
[4] N. Andhale and L. A. Bewoor, "An overview of Text Summarization techniques," 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 2016, pp. 1-7, doi: 10.1109/ICCUBEA.2016.7860024.

©IJRASET: All Rights are Reserved

[5] P. Janjanam and C. P. Reddy, "Text Summarization: An Essential Study," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi:10.1109/ICCIDS.2019.8862030.
[6] I. Awasthi, K. Gupta, Bhupendra Jogi, S. S. Anand and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1310-1317, doi:10.1109/ICICT50816.2021.9358703.
[7] H. T. Le and T. M. Le, "An approach to abstractive text summarization," 2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR), 2013, pp. 371-376, doi:10.1109/SOCPAR.2013.7054161.
[8] P. R. Dedhia, H. P. Pachgade, A. P. Malani, N. Raul and M. Naik, "Study on Abstractive Text Summarization Techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-8, doi:10.1109/ic-ETITE47903.2020.087.
[9] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)