2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)

A Hybrid Approach for Multimedia Summarization

Manikanta Avinash Grandhi, Sreebha Bhaskaran
Dept. of Computer Science & Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham, India
aviavinash1216@gmail.com, b_sreebha@blr.amrita.edu

2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon) | 978-1-6654-5342-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/NKCON56289.2022.10126694

***Abstract*——Text summarization is the basic method in NLP (Natural Language Processing) application which makes the text data into a shorter version. All over the world a lot of multimedia data is generating which includes text, audio, video, image, etc., Generating the short summary from entire text will reduce our task of understanding the things. In the proposed work, film reviews data is taken to perform summarization. A film review is an article that reviews and analyses a film and is published in a magazine, newspaper, video, or scholarly journal. As a result, reading through the entire conversation will take a long time to comprehend the reviews and consider the film. To perform text summarization, we need to work on multi-media data to generate the short summary of the entire content. By scanning the images, audio, videos and converting them into text documents, multi-media data will reduce the difficulty for the machine to understand the important information from it. In this project, proposing a Hybrid (combination of Extractive summarization and Abstractive summarization) multi-media summarization method that uses NLP techniques, image processing, speech processing to retrieve the information contained in multimedia model. The proposed work includes the sequence of implementations namely, Multimedia processing, applying NLP techniques, Seq-to-Seq modelling to the text, and performing sentiment analysis using a Naive Bayes classifier.**

***Index Terms*—Multi-Media data, Hybrid Summarization, Naive Bayes, Natural Language Processing, Sequence-to-sequencing, Paraphrasing.**

I. INTRODUCTION

In Recent years, lot of different data are generating in social media in different formats like text, audio, images, and videos which are in completely unstructured format. To store or to understand the entire text, it will take huge amount of time and space. Summarization is the technique which reduces the human work by shortening the sentences. It generates the sentences in easy and understandable way. Text Summarization the major Natural Language Processing technique, performs to summarize the text.

There are different approaches available to generate the summarized sentences like Extractive Summarization, Abstractive Summarization and Hybrid Summarization (Combination of Extractive and Abstractive summarization). In this work Hybrid Summarization is performed to get summarized data from all forms of data. The detailed explanation of work is explained in further sections as in section 2, works related and research made on work are highlighted, in section 3 the architectures used are introduced whereas in section 4 methodologies and experiments details are shown, results and implications are shown in section 5 and finally conclusion and future scope for the work are given in section 6.

II. RELATED WORKS

Exploring Sentiment Analysis on Twitter Data [1], works on different content of twitter data to perform sentiment analysis by using NLP and various ML techniques to identify the sentiment of every sentence which identifies based on the conditional probability. “Sentiment Analysis on Hindi-English Code-Mixed Social Media Text” [2] focused on identifying themixed languages of text and performing sentiment analysis us- ing attention-based CNN-Bi-LSTM model to classify the text into Positive Neutral and Negative. Kannada Speech to text Conversion using CMU Sphinx” [3], focused on converting the speech format of data into text using CMU Sphinx by training the model. CMU Sphinx supports different languages along with English. An Optimal Data Entry Method, Using Web Scraping and Text Recognition, [4] focuses on extracting the text data from the images and URL’s using OCR engine and identifying the text from it and saving it in the different formats like csv, txt, word, etc., formats “Abstractive summarization of Malayalam document using sequence to sequence model” [5] focused on the abstractive summarization of single documents in Indian regional language called Malayalam by using sequence to sequence modelling. Automated clinical concept- value pair extraction from discharge summary of pituitaryadenoma patients”.

[6] focused on to identify the summaries of discharge papers for Pituitary Adenoma which is an brain tumor disease to understand the summary by using Regular expressions and NLP rules.

The work “Read, Watch, listen and summarize” [7] is an automatic text summarization technique that takes the input from various sources and from various medias and generates the text summarization on it by taking the text dataset to train the model. This paper focused on the extractive text summarization approach by taking the small sentences from the inputs which are passed by the user itself. Regional Language Abstractive Text Summarization using Attention-based LSTM Neural Network [8] This paper is mainly on performing Abstractive approach by using two deep learning techniques called Attention based and Stacked LSTM based sequence-to-sequence neural network model. The summarization is on local regional languages of India called Hindi and Marathi languages. This summarization is of very short sentences of one to three words. “News Image captioning summarization” [9] where the summarization technique will be more useful to generate the short summary. Its main aim is to generate the short summary for the headlines or to provide some head context in the main paper by using attentional encoder decoder model by passing the news images as the input. This model is trained by taking the daily news images as the input which are created as the formatted documents

[10] In this paper, hybrid model of summarization has been implemented with Extractive and Abstractive approach by using RNN. Extractive approach uses an RNN with 2 layers and whereas the Abstractive model is being with seq2seq modelling. As Extractive approach is giving the better results, but the Abstractive approach is having less than 25% accuracy Meme-Text Analysis: Identifying Sentiment of Memes
[11] focused mostly on the sentiment analysis to find out the meaning of memes given by the memers in the social media by using Naive Bayes classification to understand the feelings and emotions

“A survey of text mining approaches, techniques, and tools on discharge summaries” [14] focused on the discharge summaries, identifies and extracts the useful information from the forms by using text mining approaches and summarizes the content into understandable format. Different ML techniques and NLP techniques has used to find out useful information from it. Automated clinical concept-value pair extraction from discharge summary of pituitary adenoma patients” “Extractive Text-Image summarization” [15] is extracting the text data from the image and summarizing the text from it by using Encoder-Decoder model with an extractive summarization approach. Daily Mail dataset is used to work on the image dataset and captions when decoding. “Multi modal Abstractive Summarization” [16] is a new multimodal summarization which uses abstractive summarization technique to summarize the 3 formats of data like audio, video, and text formats. Where abstractive will give more approximate sentences as an output by generating the new sentences from the input text with new phrases and gram mar. It shows how the method uses sequence to sequence trimodal attention to summarize the text which is an deep learning technique.
“YouTube sentiment analysis” [17] mainly focused on the sentiment analysis of the reviews particularly on the automobile reviews with the accuracy of 84%. Where the same model applied for movie reviews had given very less accurate results. It became more challenging when the model is predicting the movie re views. By this research, understood that instead of working with the reviewer’s opinion, we can focus on the reviews given by the movie lovers. When working on this data had given the better accuracy when cared with the reviewer’s opinion. For audio and video reviews, used SVM and BLSTM model to predict the sentiment analysis to estimate the accurate results.
A webpage text extraction [18] is the most important feature in multimedia summarization. In this paper, the model accepts the input as webpage URL’s or weblinks and fetches the text or images from the webpage using web scraping technique. Web scraping is mainly used for the extraction of the text from the links. Then the extracted text will be classified using the SVM and Naive Bayes Classifier. The generated output will be stored in the csv, pdf and text formats which is specified by the user. An Optimal Data Entry Method, Using Web Scraping and Text Recognition

III. SYSTEM ARCHITECTURE

This section gives a brief introduction about the architectures used in this work as follows: Hybrid Summarization is a combination of both Extractive and Abstractive approach, the input will be sent to Extractive method, where sentences will be reduced based on the scoring technique. The sentences which are having the high scoring will be sent as input to the abstractive method.

In Abstractive technique, the input will be sent to sequence-to-sequence modelling which is a deep learning technique used to reduce the sentences and generate the new sentences on its own just like a human thinking way. Fig.1 shows the complete architecture of the Hybrid summarization.

A. Extractive Summarization:

Extractive Summarization will create the summary by reusing the words, phrases, sentences, etc., from the input text. It retrieves the collection of words or sentences from the original text without modifying the document. Most of the summarization research are widely using this Extractive Summarization. Fig.2 shows the flow of execution for extractive summarization. The main disadvantage of this Extractive Summarization is it will not provide you the proper summary from the input text.

B. Abstractive Summarization:

Abstractive Summarization technique is an advanced summarization approach when compares to Extractive Summarization. It requires deep understanding of the text to generate the summary. This approach will generate the summary on its own without having the same words or sentences taken from the input text. It also determines the short, understandable, and meaning of each element like word, sentence, phrases, etc. The detailed architecture of abstractive summarization is given in Fig.3. There are lot of advantages in Hybrid Summarization, the combination of extractive and abstractive gives the better results by reducing the content and generating the better summary when compares to the individual work. Whereas the model behaves in a way how humans can think and summarize one document of text. This technique will not generate the output from the input itself, where it generates its own way of grammatical sentences.

C. Sequence-to-Sequence Modelling:

Sequence to sequence model is mostly used for the summarization. Sequence to sequence modelling with attention layer is more focusing on text features. It takes the sequence of items, words, letters, and generate another

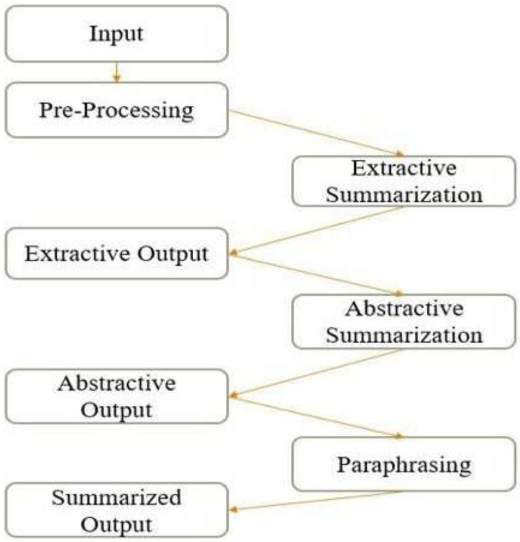

Fig. 1. Working of Hybrid Summarization.

sequence of outputs. Sequence to sequence modelling was developed by google. Its main aim is to map the fixed length of input with the fixed length of output, but in the output the length of the input and output may differ from each other. Sequential data includes text format, video format, audio format and time series data as well etc., CNN and RNN are the most important and popular algorithms that were used in sequencing models. To create a sequence-to-sequence model, there are 2 major components namely, Encoder and Decoder.

1) Encoder: In encoder LSTM is used to take the input in a sequence format. At each time stamp it will take only one character into the encoder. From the input characters, it will find the contextual information from the input sequence. The last time stamp will initialize the decoder.

2) Decoder: The decoder is also a one of the LSTM networks, will take the input from the encoder word by word and predicts the same sequence of output. This decoder is used to predict the next word of the sequence given from the previous words.

D. Naive Bayes Classifier:

The Bayes' Theorem is used to build the Naive Bayes classifiers, which are a set of classification algorithms based on the Bayes' Theorem. It's a method for categorizing text/sentences that uses supervised categorization. It's a probabilistic algorithm that assesses the likelihood of each word in a text/sentence returns the phrase with the highest likelihood. (1) is the mathematical representation of how we calculate the polarity with each word. This model, which is based on the Naive assumption that there is no relationship between independent characteristics, employs the Bayes theorem.

$$P(A|B) = P(B|A)P(A)/P(B) \tag{1}$$

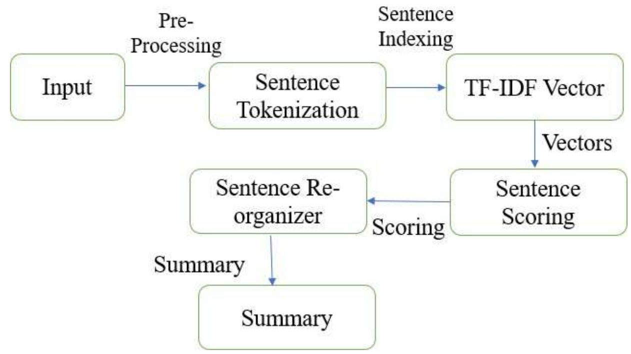

Fig. 2. Architecture of Extractive Summarization.

IV. PROPOSED METHOD

The methodology completely depends on the converting different media data to text and summarizing the sentences. This includes the steps such as collecting the data from different sources, performing pre-processing, applying different NLP techniques, building, and training the model with the train data and evaluating the model with different inputs from different web sources.

A. Data Cleaning and Preprocessing:

The data has been chosen from the Kaggle website called IMDB dataset. It has information about the Review of movie with Reference summary and rating of it. There are totally 55,50,000 reviews on different movies classified into different labels (scale of 1 to 10). There are 75% positive and 25% negative reviews available in the dataset which are equal of both the labels to train the model perfectly. The review column contains the review of the movie given by the viewers. As the reviews contains a lot of extra symbols mixed with the words,

before training the model we need to clean the data which are having null values and duplicates present in the dataset need to be removed. After cleaning the data, we need to apply data pre-processing steps to avoid the extra and unnecessary characters, symbols from the review by using NLP techniques which includes converting all characters to lower, removing html tags and extra symbols from the text, tokenizing the sentences, removing stop words, etc., from the reviews.

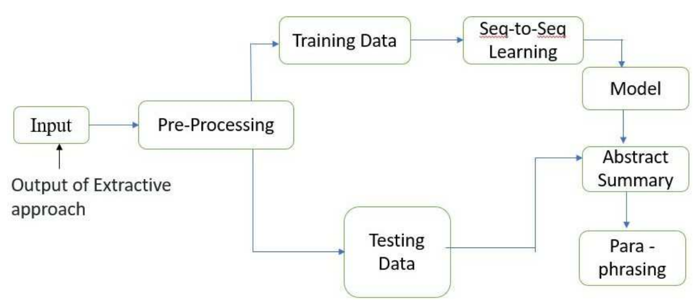

Fig. 3. Architecture of Abstractive Summarization.

B. Generating Text from Multi-Media Files:

1) Image to Text: By using Python, we can find the text data from the image and then we can convert it from image to text by using python – tesseract package which is an OCR tool for python. It recognizes the text in the images and read the text data from the given image. OCR is used to extract more than 100 languages of text from any content of document or images.

2) URL to Text: Web Scraping or web data extraction is data scraping which is used to extract the data from different websites. Beautifulsoup is one of the python libraries, used to extract the useful information from specified weblinks or websites. To perform web scrapping, we have few basic steps that we need to follow: First, need to send one HTTP request to the webpage that we want to extract the text, and that will respond with a HTML content. After sending the data, we need to fetch the data by using beautifulsoup and save the data in a dictionary or list format.

3) Audio to Text: Identifying the words and sentences from the spoken language and converting them into human understandable text is known as audio to text conversion. Speech Recognizer is the python package developed by Microsoft, used to identify the words and phrases from the audio format and convert them into text format.

4) Video to Text: To work with the video files to extract the content from it, we first need to convert the video type of data into audio format and then we can easily extract the text type of data from the audio format by using the same package called Speech Recognizer which is an Microsoft service to extract the useful content from the video input. To convert from video format to audio format, we use “audio.write audiofile” to convert into audio “.wav” format.

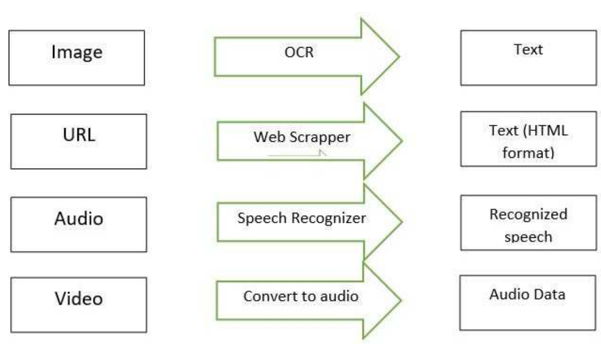

Fig. 4. Architecture of Text to Multi-Media Files.

C. Hybrid Text Summarization:

After cleaning and preprocessing the text input, the data is ready to apply the summarization techniques. The cleaned text will be taken and applied with tokenization and stop words. Tokenization is the process of dividing the sentences into individual words. Stop words are the common words to ignore the common words which doesn’t have any values unless to maintain as grammatical sentences. Initially, input will be divided into different sentences and indexed based on the order. After that, by tokenizing the input, each sentence will be divided into individual words. For each word in the input, applying some feature extraction to convert the text into vector representation.

1) Extractive Approach: TF-IDF: method used to generate the vector representation of input. TFIDF is the Term Frequency – Inverse Document Frequency, creates the semantic representation for the texts in matrix format. Where term-frequency will be calculated in such a way that the total number of times particular word has occurred will be divided with the total words in the document or in sentence. equation 3 shows the mathematical representation of Term-Frequency. The Inverse Document Frequency will be calculated as how many documents that word present in it will be divided with total number of documents and applied logarithm to it. equation 4 shows the mathematical representation of Inverse Document Frequency. TF-IDF is more capable of handling the large amount of data and documents. As it is one of the best approaches to generate vector representations. The equation 2 is the mathematical representation of how each word will be converted into numerical representation.

Sentence Scoring: After getting the vector representation for each word in the input, total words in sentence will be summed up and added as one more column called sentence scoring. But there might be a chance of more sparse values in the representation. So, to avoid the sparse values, taking only the top-n/2 words in each sentence and added up and stored as sentence scoring. Based on the scoring of sentences, Top-K sentences will be taken into consideration and then sentences will be re-organized based on the indexing and generated as output. This approach of reducing the unwanted text from the input can be called as Extractive Summarization. Now, whatever the output that we got from extractive summarization, that will be taken as input to the abstractive summarization.

2) Abstractive Approach: Here the abstractive summarization will use the deep learning approach called sequence-to-sequence modelling to build the model that will generate the final summarization part with new grammatical sentences based on the importance of the words. To work with abstractive approach, we need a dataset to train the model to predict the results on its own. The IMDB dataset will be used to train the deep learning model called Sequence-to-Sequence modelling with attention layer which focuses more on text features. It takes the input as sequence of words and generates another sequence of outputs. The sentences will be turned into three NumPy arrays, which are encoder input data, decoder input data, and decoder target data. Then, need to train a sequence-to-sequence with LSTM based model to predict the target data of decoder which the input will be encoder and decoder input data. Later on, the test data will be used to test the model to check the accuracy of the results.

Sequence-to-Sequence with attention mechanism:

First, we need to define input sequence and process it, setup the decoder as initial state by using encoder states. Create the model afterwards so that it can generate the decoder target data from the encoder and decoder input data. Using the test data, train the model and make predictions. Encode the input as state vectors and create an empty target sequence of length one in order to decode the vector sequence. The first character of the target sequence should be filled with the start character. Sample loop for a group of sequences with the following exit criteria: either reach maximum length or locate stop character. Update the target sequence (of length 1) and Update states returndecoded sentences We use each time step of the encoder to implement the attention mechanism, but we give the time steps more weight. The weight relies on how crucial that time step was for the decoder to produce the following word in the sequence as ideally as possible. Fig 5 will show you the training accuracy and loss of the model, the graph drawn for 500 epochs with 5000 records of data. When the number of epochs increases, accuracy will be increased, and loss will be increased.

Mathematical approach of how TF-IDF works is shown in (2),

$$tfidf(t, d) = tf(t, d) log(N/(df + 1)) \quad (2)$$

Where TF is Term Frequency,

$$tf(t, d) = count\ of\ t\ in d/number\ of\ words\ in d \quad (3)$$

IDF is Inverse Document Frequency,

$$idf(t) = N/df \quad (4)$$

D. sentiment Analysis using Naive Bayes Classifier:

Sentiment Analysis is the classification of the data based on the input given by the user. It is an automatic process which identifies the customers opinion and categorizing them as positive, neutral, and negative based on the customers expression on the product. Sentiment Analysis uses different Data Mining methods to extract and clean the data by using different Natural Language Processing techniques. Basically, sentiment analysis is finding the polarity of the word

data and labelling them into positive, negative, and neutral. Before classifying the summary of the movie reviews, we need to apply one of the popular techniques called “Bag of Words” (Bow), which converts the documents or the input into vectors and then the vectors. Here each word in the input will be assigned with some score. Fig shows the complete process of unstructured data to sentiment labelling of data.

A set of straightforward probabilistic classifiers called Multinomial Naive Bayes (MNB) classifiers are based on the widely accepted idea that, given the category variable, all features are independent of one another. The MNB is used due to its simplicity in the training and classification phase. Naive Bayes Requires a small amount of training data to learn the parameters.

Fig. 5. Training Accuracy and Loss of Model.

V. RESULTS & DISCUSSION

In this work, IMDB dataset has taken to train and test the models. The dataset is taken from the open-source website Kaggle.com. It consists of 50k rows which includes the details about the movie like MovieID, MovieName, rating, review, reviewSummary, reviewDate, sentiment analysis and etc., As the work is completely on summarizing part, only review and reviewSummary has taken into consideration to train and test the model. In the dataset, there are 999 movie names are included which are unique. The reviewers have given reviews to all the movies. It consists of different movie reviews along with referenceSummary of it. To train the sequence-to-sequence model with attention layer data splitted into train and test parts. Train data will be used give training to the model, to understand the data for the model. Whereas the model will be tested with the test data to check how better the model understands and trained itself to get the accurate results.

The output of the model will be differentiated with the reference summary based on the evaluation metric called Rogue, used for the evaluation of automatic text summarization methods. Evaluating the results with unigram and bigram and longest sequence called Rouge-L to evaluate the summaries in terms of precision, recall and F1 score.

In TABLEI, we have a result of the model separately to check the recall, precision of the model. Where we can see the results are giving more accurate information when compared to the base model that can achieve competitive results.

TABLE I
RESULTS OF OUR APPROACH

		Precision	Recall	F1
Extractive summarization	R1	36.9	59.26	45.48
	R2	18.02	22.63	20.06
	R-L	34.21	50.82	40.89
Abstractive summarization	R1	42.36	48.24	45.10
	R2	21.5	26.08	23.56
	R-L	42.23	43.97	43.08
Hybrid summarization	R1	39.42	55.60	46.13
	R2	16.72	31.54	21.85
	R-L	38.23	47.65	42.42

TABLE II
COMPARISON OF PROPOSED APPROACH WITH EXISTING WORK

		Precision	Recall	F1
Extractive	SummeRuNNEr [12]	39.6	16.2	35.3
	Our approach	45.48	20.06	40.89
Abstractive	[13]	38.1	13.9	34.0
	Our approach	45.1	23.56	43.08
Hybrid	Our approach	46.13	21.85	42.42

In TABLEII, we have 3 blocks, includes extractive, abstractive and hybrid model results. First block contains extractive model results that reaches respectively, 45.48, 20.06, 40.89 for R-1, R-2 and R-L values. Whereas 39.6, 16.2 and 35.3 of R-1, R-2 and R-L values for the base model called SummeRuNNEr. The base model mainly focused on the first three sentences with the high F-1 score. Whereas our model focused on the adjusting of the summary length based on the probability. In the second block, abstractive results are trained, as our approach reaches 45.1 for R-1, 23.56 for R-2 and 43.08 of R-L which is better than the base model, as our approach improvesthe attention mechanism. The third block of table -2 containsthe results of the hybrid approach which gave the better of 46.13 for R-1, 21.85 for R-2 and 42.42 of R-L values. As it is giving the better results when combining both the extraction and abstraction models.

VI. CONCLUSION & FUTURE SCOPE

In this paper, we proposed the text summarization method. Different approaches have been followed and evaluated with Rouge Metrics like developing the extractive approach by using vector representation and scoring techniques, and developed the abstractive approach with sequence-to-sequence modelling with attention mechanism. Extractive approach is the best part of producing the good abstractive results. First results of extractive and abstractive approaches gave the confidence to the hybrid approach which provides the good qualityof summaries.

As a future scope, we can also integrate the model with transformer layers to check how the attention mechanism works. Also, we can improve the accuracy of the model by adding the techniques of reinforcement learning and using BERT for language generation. Also, as a extension, we can also train the model with different regional and foreign languages to summarize the text.

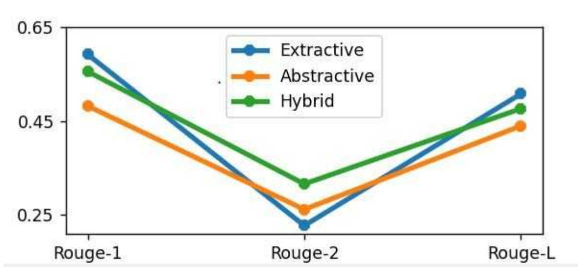

Fig. 6. Comparison chart for average recall

REFERENCES

[1] Venugopalan, Manju, and Deepa Gupta.” Exploring sentiment analysis on twitter data.” In 2015 eighth international conference on contemporary computing (IC3), pp. 241-247. IEEE, 2019.
[2] Tulasi Sasidhar, T., B. Premjith, K. Sreelakshmi, and K. P. Soman.” Sentiment Analysis on Hindi–English Code-Mixed social media Text.” In Innovations in Computer Science and Engineering, pp. 615-622. Springer, Singapore, 2021.
[3] Shivakumar, K. M., K. G. Aravind, T. V. Anoop, and Deepa Gupta.” Kannada speech to text conversion using CMU Sphinx.” In 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 3, pp. 1-6. IEEE, 2016.
[4] Roopesh, N., M. S. Akarsh, and C. Narendra Babu.” An Optimal Data Entry Method, Using Web Scraping and Text Recognition.” In 2021 International Conference on Information Technology (ICIT), pp. 92-97. IEEE, 2021.
[5] Nambiar, Sindhya K., S. David Peter, and Sumam Mary Idicula. ”Abstractive Summarization of Malayalam Document using Sequence to Sequence Model.” In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 347-352. IEEE, 2021. Harvard
[6] Nair, Priyanka C., Deepa Gupta, Bhagavatula Indira Devi, and Nikhil Raghav Bhat.” Automated clinical concept-value pair extraction from discharge summary of pituitary adenoma patients.” In 2019 9th International Conference on Advances in Computing and Communication (ICACC), pp. 258-264. IEEE, 2019.
[7] Li, Haoran, Junnan Zhu, Cong Ma, Jiajun Zhang, and Chengqing Zong. ” Read, watch, listen, and summarize: Multi-modal summarization for asynchronous text, image, audio and video.” IEEE Transactions on Knowledge and Data Engineering 31, no. 5 (2018): 996-1009.
[8] Karmakar, Rishabh, Ketki Nirantar, Prathamesh Kurunkar, Pooja Hiremath, and Deptii Chaudhari.” Indian Regional Language Abstractive Text Summarization using Attention-based LSTM Neural Network.” In 2021 International Conference on Intelligent Technologies (CONIT), pp. 1-8. IEEE, 2021.
[9] Chen, Jingqiang, and Hai Zhuge.” News image captioning based on text summarization using image as query.” In 2019 15th International Conference on Semantics, Knowledge, and Grids (SKG), pp. 123-126. IEEE, 2019.
[10] Chang, Te-Min, and Wen-Feng Hsiao.” A hybrid approach to automatic text summarization.” In 2008 8th IEEE International Conference on Computer and Information Technology, pp. 65-70. IEEE, 2008.
[11] P Aditya Krishna Rohith, Sachin Kumar S, K P Soman,” Meme-Text Analysis: Identifying Sentiment of Memes”, ADSC, 2022
[12] Tan, Jiwei, Xiaojun Wan, and Jianguo Xiao.” Abstractive document summarization with a graph-based attentional neural model.” In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1171-1181. 2017.
[13] Nallapati, Ramesh, Feifei Zhai, and Bowen Zhou.” Summarunner: A recurrent neural network-based sequence model for extractive summarization of documents.” In Thirty-first AAAI conference on artificial intelligence. 2017.
[14] Nair, Priyanka C., Deepa Gupta, and Bhagavatula Indira Devi.” A survey of text mining approaches, techniques, and tools on discharge summaries.” In Advances in Computational Intelligence and CommunicationTechnology, pp. 331-348. Springer, Singapore, 2021.
[15] Chen, Jingqiang, and Hai Zhuge.” Extractive text-image summarization using multi-modal RNN.” In 2018 14th International Conference on Semantics, Knowledge, and Grids (SKG), pp. 245-248. IEEE, 2018.
[16] Palaskar, Shruti, Jindrich Libovický, Spandana Gella, and Florian Metze. ”Multimodal abstractive summarization for how2 videos.” arXiv preprint arXiv:1906.07901 (2019).
[17] Wöllmer, Martin, Felix Weninger, Tobias Knaup, Björn Schuller, Congkai Sun, Kenji Sagae, and Louis-Philippe Morency. ” YouTube movie reviews: Sentiment analysis in an audio-visual context.” IEEE Intelligent Systems 28, no. 3 (2013): 46-53.
[18] Wang, Chongjun, and Peng Wei. ”A novel web page text information extraction method.” In 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 2213-2218. IEEE, 2019.