SURVEY OF VARIOUS MACHINE LEARNING ALGORITHMS FOR TEXT SUMMARIZATION OF ONLINE NEWS ARTICLE

[1]**Arokiaraj Christian Saint Hubert, **[2]**Kiran Gowtham S VS, **[3]**Prabalesh, **[4]**Roger Louis**

[1]Assistant Professor, [2]Student, [3]Student, [4]Student

[1]Department Computer Science and Engineering

[1]Sri Manakula Vinayagar Engineering College, Puducherry, India

Abstract: In the contemporary world, the traditional practice of reading entire newspapers daily is challenging, given the plethora of sources and potential information adulteration due to political biases. This work proposes an automated news article summarization system employing Natural Language Processing (NLP), a subset of Artificial Intelligence (AI), to condense lengthy articles and save readers valuable time. With over 1,11,000 registered newspapers in India, the system aims to filter legitimate information from various online sources, overcoming issues such as bias and human error [7]. The proposed architecture involves fetching the top ten news articles from Google News via its API, utilizing Python's Beautiful Soup for web scraping, and employing Bidirectional Long Short-Term Memory (BiLSTM) to assess information accuracy across multiple articles. The resulting summaries provide readers with a comprehensive understanding of the news without the need to read the entire article, offering an efficient and time-saving approach to news consumption in the digital age.

Keywords - *Bidirectional Long Short-Term Memory (BiLSTM), employing Natural Language Processing (NLP), a subset of Artificial Intelligence (AI).*

I. INTRODUCTION

The news serves as information about current events and can be disseminated through various mediums such as word of mouth, printing, postal systems, broadcasting, and electronic communication. It encompasses a wide range of topics, including war, government, politics, education, health, the environment, economy, business, fashion, entertainment, and sports, as well as unusual or quirky events. The term "hard news" distinguishes factual reporting from softer media content. Historical practices of conveying news orally evolved into the establishment of newspapers in Europe during the early modern period, having originated in China over centuries. The 20th century saw radio and television emerge as significant news transmission channels, and in the 21st century, the internet has assumed a pivotal role in rapidly spreading and influencing the content of news, marking a transformative shift in how information is shared and consumed.

In the era of big data, the abundance of text data from various sources has surged, necessitating effective summarization to extract valuable information and knowledge. Natural Language Processing (NLP) plays a crucial role in addressing this challenge through automatic text summarization. This process involves distilling a comprehensive and coherent summary from a large volume of literature without human intervention while preserving the essence of the original text. Unlike humans who read and comprehend text before summarizing, computers lack inherent language understanding, making automatic text summarization a complex and time-intensive task. The goal is to capture the most pertinent elements of a document, providing a condensed highlight of its content. This is vital for efficiency, enabling users to quickly consume material while saving time in the face of the overwhelming volume of available data.

1.1 DOMAIN INTRODUCTION

Deep learning is a subfield of machine learning that leverages artificial neural networks to simulate the human brain's structure and function in order to analyze and process data. These neural networks, composed of interconnected layers of nodes or neurons, can automatically learn hierarchical representations from the input data. Deep learning has demonstrated remarkable success in various tasks, such as image and speech recognition, natural language processing, and decision-making. The depth of these neural networks enables them to discern intricate patterns and features, making them particularly effective in handling large and complex datasets. Training deep learning models involves adjusting the weights and biases of the connections between neurons based on the error generated during predictions, allowing the system to iteratively refine its performance and enhance its ability to generalize to new, unseen data.

Deep learning algorithms have fueled advancements in diverse domains, from computer vision and speech processing to autonomous vehicles and healthcare diagnostics. The transformative power of deep learning lies in its capacity to automatically extract relevant features from raw

data, eliminating the need for manual feature engineering. This adaptability and scalability make deep learning a versatile and potent tool for solving complex problems across various disciplines.

1.2 NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) plays a pivotal role in the realm of news summarization, leveraging advanced algorithms and linguistic analysis to automatically process and understand human language. In the context of news summarization, NLP allows systems to sift through vast amounts of textual information from multiple sources, extracting key insights, identifying significant entities, and discerning the overall sentiment of the content. This capability is crucial for distilling the most relevant and salient information from diverse news articles, enabling the creation of concise and coherent summaries.

In the news summarization process, NLP techniques, such as named entity recognition, sentiment analysis, and part-of-speech tagging, contribute to the extraction of essential details from articles. Additionally, the application of NLP helps in overcoming language nuances and complexities, ensuring that the generated summaries maintain coherence and readability. By automating the extraction and condensation of information, NLP not only enhances the efficiency of news consumption but also addresses the challenges posed by the sheer volume of data available in the digital age, providing users with succinct and informative summaries of current events.

1.3 APPLICATION OF THE PROJECT

The application of news summarization holds significant value across various domains, offering practical benefits in information retrieval, time efficiency, and decision-making processes. In the realm of journalism and media, automated news summarization systems powered by Natural Language Processing (NLP) enable news agencies to quickly process vast amounts of information from diverse sources. By condensing articles into succinct summaries, journalists can stay updated on multiple stories efficiently, facilitating faster news production cycles. Beyond media, news summarization finds applications in business intelligence, research, and information management. Professionals in these fields can utilize automated summarization tools to quickly grasp the key points of news articles relevant to their industry or research focus. Utility from journalism to various sectors, enhancing information accessibility and decision-making processes.

1.4 ALGORITHM USED

The Bidirectional Long Short-Term Memory (BiLSTM) algorithm is an extension of the traditional Long Short-Term Memory (LSTM) model, designed to enhance the understanding of sequential data. In a BiLSTM network, information is processed not only in a forward manner but also in a backward direction, enabling the model to capture dependencies from both past and future context.
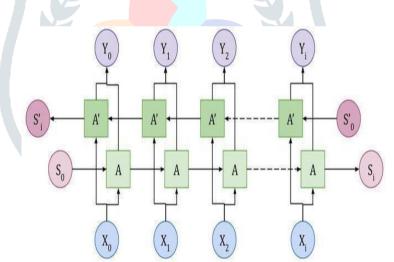

Figure. 1 Bidirectional LSTM layer Architecture

This bidirectional processing is particularly useful in tasks where context from both preceding and succeeding elements in a sequence is crucial, such as natural language processing and time-series analysis. The bidirectional nature of the BiLSTM allows it to capture long-term dependencies more effectively, making it well-suited for tasks involving sequential data where a comprehensive understanding of context is essential.

1.5 ADVANTAGES OF BI-LSTM ALGORITHM

The Bidirectional Long Short-Term Memory (BiLSTM) algorithm offers several advantages, particularly in the context of sequence modelling. This bidirectional approach enhances the model's ability to grasp intricate patterns and relationships within sequential data, leading to improved performance in tasks that require a nuanced understanding of context. Overall, the BiLSTM algorithm's advantage lies in its capacity to enhance the modelling of sequential data by considering information from both historical and future perspectives, contributing to more robust and accurate predictions or classifications.

II. REVIEW OF LITERATURE

2.1   Automatic News Article Summarization by Laxmi B. Rananavare, P. Venkata Subba Reddy

A summary condenses a lengthy document by highlighting salient features. It helps the reader to understand completely just by reading the summary so that the reader can save time and also decide whether to go through the entire document. Summaries should be shorter than the original article so make sure to select only pertinent information to include the article. The main goal of a newspaper article summary is, for the readers to walk away with knowledge of what the newspaper article is all about without the need to read the entire article. This work proposes a news article summarization system which accesses information from various local online newspapers automatically and summarizes information using heterogeneous articles. To make ad-hoc keyword-based extraction of news articles, the system uses a tailor-made web crawler which crawls the websites for searching relevant articles. Computational Linguistic techniques mainly Triplet Extraction, Semantic Similarity calculation and OPTICS clustering with DBSCAN are used alongside a sentence selection heuristic to generate coherent and cogent summaries irrespective of the number of articles supplied to the engine. The performance evaluation is done using the ROUGE metric.

2.2 News Article Summarization with Attention-based Deep Recurrent Neural Networks by Hujia Yu, Chang Yue

A news article often has a brief title that summarizes its content for readers to decide whether they want to read further. Therefore, a well-summarized title is crucial in delivering the main point of the news article to its potential readers. As a result, automatic text summarization techniques have a huge potential for news articles in that they expedite the process of summarizing given documents for humans and, if models are well-trained, generate the summary with high accuracy. Our goal is to build a text automatic summarization model with deep learning algorithms that can output a one-sentence summarization given an article. Models developed for text auto summarization have immediate applications in news article title generations and beyond, such as machine translation, image captioning, as well as video summarization. There are in general two types of summarization techniques: extractive summarization and abstractive summarization, where the former summarizes articles by selecting a subset of words that remain. We approached the problem with an attentive recurrent neural network model. We further explored the encoder-decoder model with GRU and LSTM cells. In addition, we also briefly investigated models with stacked RNNs. We wanted to see whether differing layers of RNNs can capture semantic meanings between words as we intended. Our dataset is 'The Signal Media One-Million News Articles Dataset', which contains a million news articles and their titles in the form of a JSON file.

2.3 Abstractive Sentence Summarization with Attentive Recurrent Neural Networks by Sumit Chopra, Michael Auli

Abstractive Sentence Summarization generates a shorter version of a given sentence while attempting to preserve its meaning. We introduce a conditional recurrent neural network (RNN) which generates a summary of an input sentence. The conditioning is provided by a novel convolutional attention-based encoder which ensures that the decoder focuses on the appropriate input words at each step of generation. Our model relies only on learned features and is easy to train in an end-to-end fashion on large data sets. Our experiments show that the model significantly outperforms the recently proposed state-of-the-art method on the Gigaword corpus while performing competitively on the DUC-2004 shared task. Generating a condensed version of a passage while preserving its meaning is known as text summarization. Tackling this task is an important step towards natural language understanding. Summarization systems can be broadly classified into two categories. Extractive models generate summaries by cropping important segments from the original text and putting them together to form a coherent summary. Abstractive models generate summaries from scratch without being constrained to reuse phrases from the original text. The main contribution of this paper is a novel convolutional attention-based conditional recurrent neural network model for the problem of abstractive sentence summarization. Empirically we show that our model beats the state-of-the-art systems of Rush et al. (2015) on multiple data sets. Particularly notable is the fact that even with a simple generation module, which does not use any extractive feature tuning, our model manages to significantly outperform their ABS+ system on the Gigaword data set and is comparable on the DUC-2004 task.

2.4 A Review of Natural Language Processing Research by Lene Casey, Emma Davidson, Michael Poon, Hang Dong.

Natural language processing (NLP) is a theory-motivated range of computational techniques for the automatic analysis and representation of human language. NLP research has evolved from the era of punch cards and batch processing (in which the analysis of a sentence could take up to 7 minutes) to the era of Google and the likes it (in which millions of web pages can be processed in less than a second). We present a comprehensive analysis of the 164 publications retrieved with publications in 2019 almost triple those in 2015. Each publication is categorized into one of 6 clinical application categories. Deep learning use has increased in this period but conventional machine learning approaches are still prevalent. Deep learning remains challenged when data is scarce and there is little evidence of adoption into clinical practice. Despite 17% of studies reporting greater than 0.85 F1 scores, it is hard to comparatively evaluate these approaches given that most of them use different datasets. Only 14 studies made their data and 15 their code available with 10 externally validating results.

2.5 Review of Automatic Text Summarization Techniques by Adhika Pramita Widyassari, Supriadi Rustad

Text summarization automatically produces a summary containing important sentences and includes all relevant important information from the original document. One of the main approaches, when viewed from the summary results, is extractive and abstractive. An extractive summary is heading towards maturity and now research has shifted towards abstractive summation and real-time summarization. Although there have been so many achievements in the acquisition of datasets, methods, and techniques published, there are not many papers that can provide a broad picture of the current state of research in this field. This paper provides a broad and systematic review of research in the field of text summarization published from 2008 to 2019. There are 85 journal and conference publications which are the results of the extraction of selected studies for identification and analysis to describe research topics/trends, datasets, preprocessing, features, techniques, methods, evaluations, and problems in this field of research. The results of the analysis provide an in-depth explanation of the topics/trends that are the focus of their research in the field of text summarization; provide references to public datasets, preprocessing and features that have been used; describe the techniques and methods that are often used by researchers as a comparison and means for developing methods. At the end of this paper, several recommendations for opportunities and challenges related to text summarization research are mentioned.

2.6 Automatic lossless-summarization of news articles with abstract meaning representation by Ritwik Mishra, Tirthankar Gayenb

The continuously growing size of the textual content needs a way to be designed to store the information in a condensed manner with minimal information loss. Hence, the concept of lossless-summary is introduced which aims to address the problem of dangling anaphoras and incoherency in extractive summaries. A pipeline of operations has been proposed by this work in order to generate lossless summaries. Coreference resolution is performed pairwise on the sentences before generating the Abstract Meaning Representation (AMR) of the sentences. An algorithm to merge AMR graphs is developed and finally, the text is generated using the merged AMR graphs. CNN/Dailymail dataset of news articles is used for evaluations and results obtained show the potential for lossless summarisation. The task of summarization can be defined by creating a new condensed version of the reference text that is smaller in size but retains various characteristics of the original document. A document summary can be of two types: (a) Extractive: where the relevant sentences or part of sentences are extracted verbatim from the reference text, and (b) Abstractive: where the important information in reference text is paraphrased and rewritten. The field of automatic text summarization was initiated in the late 1950s by HP Luhn 3. It was an extractive approach which used features such as word frequency to rank the sentences. Literature suggests that since then, most of the work in automatic summarization deals with generating extractive summaries.

2.7 Rapid Detection of Fake News based on Machine Learning methods by Barbara Probierz, Piotr Stefański, Jan Kozak

Nowadays, it is very important to quickly recognize false information referred to as fake news. This is especially important in the case of news appearing on the Internet because of its wide and rapid spreading. It is equally important to be able to initially classify news as fake or true based on the title itself. In this paper, we propose an approach to classifying news based on the title without analyzing the other aspects. The obtained results will be compared with classification based on the whole news text. The goal of this work is to propose a method that balances data analysis time and quality of classification in fake news prediction. We use natural language processing (NLP) to describe the title and text of the news. This is a complex process, requiring good analysis to be applied to classification. Therefore, the use of complex classifiers – in this case, classical ensemble methods – has been proposed in order to achieve a high quality of classification (measured by popular measure). In this paper, we present analyses of a real data set and results of news classification using the proposed model – including an ensemble of classifiers and single classifiers.

2.8 Fake News Detection Using Machine Learning Ensemble Methods by Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf and Muhammad Ovais Ahmad

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in human history before. With the current usage of social media platforms, consumers are creating and sharing more information than ever before, some of which are misleading with no relevance to reality. Automated classification of a text article as misinformation or disinformation is a challenging task. Even an expert in a particular domain has to explore multiple aspects before giving a verdict on the truthfulness of an article. In this work, we propose to use a machine learning ensemble approach for the automated classification of news articles. Our study explores different textual properties that can be used to distinguish fake content from real. By using those properties, we train a combination of different machine learning algorithms using various ensemble methods and evaluate their performance on 4 real-world datasets. Experimental evaluation confirms the superior performance of our proposed ensemble learner approach in comparison to individual learners. The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in human history before. Besides other use cases, news outlets benefitted from the widespread use of social media platforms by providing updated news in near real-time to their subscribers. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats [17]. It became easier for consumers to acquire the latest news at their fingertips. Facebook referrals account for 70% of traffic to news websites [7]. These social media platforms in their current state are extremely powerful and useful for their ability to allow users to discuss and share ideas and debate over issues such as democracy, education, and health. However, such platforms are also used with a negative perspective by certain entities commonly for monetary gain [2, 8] and in other cases for creating biased opinions, manipulating mindsets, and spreading satire or absurdity. The phenomenon is commonly known as fake news. There has been a rapid increase in the spread of fake news in the last decade, most prominently observed in the 2016 US elections [6].

2.9 Grouping sentences as better language unit for extractive text summarization by Mengyun Cao, Hai Zhuge

Most existing methods for extractive text summarization aim to extract important sentences with statistical or linguistic techniques and concatenate these sentences as a summary. However, the extracted sentences are usually incoherent. The problem becomes worse when the source text and the summary are long and based on logical reasoning. The motivation of this paper is to answer the following two related questions: What is the best language unit for constructing a summary that is coherent and understandable? Extracting larger language units such as a group of sentences or a paragraph is a natural way to improve the readability of the summary as it is rational to assume that the original sentences within a larger language unit are coherent. This paper proposes a framework for group-based text summarization that clusters semantically related sentences into groups based on Semantic Link Network (SLN) and then ranks the groups and concatenates the top-ranked ones into a summary. A two-layer SLN model is used to generate and rank groups with semantic links including the is-part-of link, sequential link, similar-to link, and cause–effect link. The experimental results show that summaries composed by group or paragraph tend to contain more keywords or phrases than summaries composed by sentences and summaries composed by groups contain more keywords or phrases than those composed by paragraphs especially when the average length of source texts is from 7000 words to 17,000 words which is the usual length of scientific papers. Further, we compare seven clustering algorithms for generating groups and propose five strategies for generating groups with the four types of semantic links

2.10 Rapid detection of fake news based on machine learning methods by Arbara Probierz, Piotr Stefański, Jan Kozak

Nowadays, it is very important to quickly recognize false information referred to as fake news. This is especially important in the case of news appearing on the Internet because of its wide and rapid spreading. It is equally important to be able to initially classify news as fake or true based on the title itself. In this paper, we propose an approach to classifying news based on the title without analyzing the other aspects. The obtained results will be compared with classification based on the whole news text. The goal of this work is to propose a method that balances data analysis time and quality of classification in fake news prediction. We use natural language processing (NLP) to describe the

title and text of the news. This is a complex process, requiring good analysis to be applied to classification. Therefore, the use of complex classifiers – in this case, classical ensemble methods – has been proposed in order to achieve a high quality of classification (measured by popular measure). In this paper, we present analyses of a real data set and results of news classification using the proposed model – including an ensemble of classifiers and single classifiers.

III. PROBLEM IDENTIFICATION

The identified problem in the project stems from the impracticality of reading entire newspapers daily in the modern era, exacerbated by the sheer volume of registered newspapers in India, amounting to 1,10,851. The challenge is twofold – distinguishing legitimate information from the few reliable sources amidst the plethora of newspapers and addressing the issue of potential adulteration of information driven by political motives. Additionally, the presence of human error in newspaper content further complicates the quest for accurate and unbiased news consumption.

To tackle this problem, the project proposes the utilization of Natural Language Processing (NLP), a subdomain of Artificial Intelligence (AI), to create concise and accurate summaries from multiple newspaper sources.

The focus is on keyword extraction through NLP to identify key topics, and the verification of these keywords using the Long Short-Term Memory (LSTM) algorithm. The ultimate goal is to generate brief and precise summaries that provide the essence of news articles, allowing individuals to stay informed without the need to sift through extensive content.

This approach aims to streamline news consumption, combat information overload, and enhance the accuracy of information accessed by individuals in the midst of the vast and potentially biased landscape of news sources.

IV. PROBLEM DEFINITION

The problem addressed by the project lies in the contemporary challenge of navigating through an overwhelming volume of newspapers in India, totalling 1,10,851, and discerning reliable information amidst potential political bias and human errors. In the fast-paced modern lifestyle, reading entire newspapers daily has become impractical, necessitating a more efficient way to consume news. The proposed solution leverages Natural Language Processing (NLP), a subset of Artificial Intelligence, to extract key information and generate succinct summaries from multiple newspaper sources.

The use of the Long Short-Term Memory (LSTM) algorithm further ensures the accuracy of the summarized data, allowing individuals to stay well-informed with concise statements that capture the essence of news articles. This project addresses the need for an automated, reliable, and efficient method of news summarization to facilitate informed decision-making and mitigate the challenges posed by the vast and potentially biased landscape of news media.
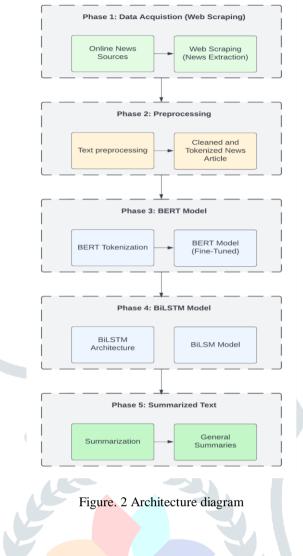
4.1 ARCHITECTURE DIAGRAM

The architecture for news summarization represents a cutting-edge fusion of technology and linguistics, finely tuned to navigate the ever-expanding landscape of digital news content. It operates through a meticulously orchestrated trio of components, each playing a pivotal role in the synthesis of insightful summaries. At its inception, the system embarks on a quest for knowledge, utilizing topic-based article retrieval mechanisms to sift through a labyrinth of news sources. This initial phase is akin to casting a wide net, casting across the digital seas to capture articles relevant to the user's specified topic. Through sophisticated algorithms, the system identifies and collects a diverse array of articles, ensuring a comprehensive pool from which to draw insights.

As the retrieved articles flood in, the Natural Language Processing (NLP) engine takes centre stage, wielding the power of computational linguistics to unravel the complexities of human language. Within its digital domain, NLP meticulously dissects each article, extracting nuggets of information with surgical precision. Named entities emerge from the textual tapestry, sentiments are quantified, and linguistic nuances are decoded, painting a rich and textured portrait of the news landscape. But the true magic unfolds in the final act, where the summarization generation module transforms raw data into elegant prose. Armed with the distilled essence of countless articles, this module weaves together a narrative tapestry, distilling the abundance of information into a concise and coherent summary. Through a delicate balance of brevity and depth, the summary emerges as a beacon of clarity amidst the sea of digital noise, offering users a lucid window into the heart of the news.
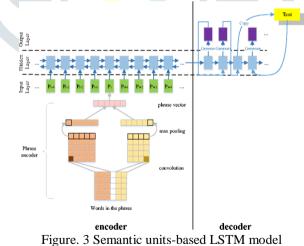
In its essence, the streamlined architecture for news summarization emerges as a beacon of efficiency and accuracy amidst the tumultuous seas of digital media. Through seamless integration, it orchestrates a symphony of information retrieval, content analysis, and summarization generation, empowering users with a reliable compass through the labyrinthine currents of current events. This paradigm shift in information engagement replaces the days of drowning in unfiltered data with sophisticated tools capable of distilling complex news topics into easily digestible snippets. In an era defined by information overload, this architecture offers respite by parsing vast data, extracting insights, and presenting them coherently, reshaping our understanding of the world and empowering informed decision-making in the complex global landscape.

Figure. 2 Architecture diagram

4.2 COLLECTION OF THE KEYWORDS USING THE NLP

In this module, the collected data is transformed into a pandas data frame for efficient processing. Natural Language Processing (NLP) techniques, specifically sentiment analysis, are applied to analyze the textual content. Keywords are identified and removed from the data, and a scoring mechanism evaluates the significance of each keyword. Subsequently, the data is summarized based on the assessed scores, and the condensed information is stored. This streamlined process combines data manipulation, sentiment analysis, and keyword scoring to produce a concise and informative summary of the collected data.
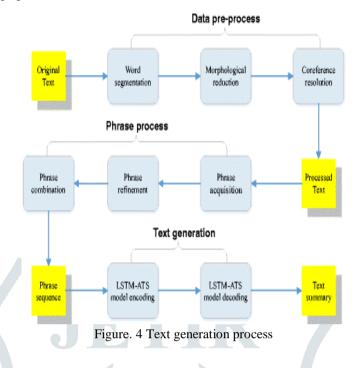

Figure. 3 Semantic units-based LSTM model

4.3 PREDICTION OF ACCURACY USING THE BI-LSTM ALGORITHM

In this module, the collected data from websites undergoes a two-step verification process. Initially, Natural Language Processing (NLP) with sentiment analysis is applied to separate keywords from the text. These keywords serve as crucial elements for summarization. Subsequently, the accuracy of these extracted keywords is validated using the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm. This bidirectional approach enhances the precision of keyword identification. Finally, the news summaries obtained from the websites are cross-verified, ensuring that the generated summaries align accurately with the validated keywords. This dual verification process enhances the reliability and accuracy of the news summarization system.

4.4 SUMMARIZING THE NEWS DATA

The collected data undergoes prediction using the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm, enabling the system to discern patterns and extract meaningful insights. Subsequently, the predicted data is summarized using the Bi-LSTM, condensing the essential information. The summarized news, sourced from various websites, undergoes verification to ensure accuracy and reliability. This streamlined process employs advanced machine learning techniques to distil relevant information, providing users with a succinct and verified overview of significant news in a concise paragraph.

Figure. 4 Text generation process

V. CONCLUSION

In conclusion, our proposed system centers on developing a text-automatic summarization model tailored for news articles, with the primary objective of generating concise one-sentence summaries that mirror the style of news titles. The successful integration of the news summarizer, employing the Bidirectional Long Short-Term Memory (Bi-LSTM) deep learning algorithm in conjunction with Natural Language Processing (NLP), underscores our commitment to harnessing advanced technologies in the service of efficient information dissemination. Our recognition of sentiment analysis as a potent method for text summarization underscores our intention to pursue this avenue in future iterations, leveraging nuanced insights to enrich the summarization process and enhance user engagement.

VI. FUTURE IMPLEMENTATION

Acknowledging the existing limitations of our model, we are dedicated to refining our system through a multipronged approach. This includes refining parameters through enhanced fine-tuning, exploring novel deep learning methodologies to further optimize performance, and augmenting the training dataset to ensure robustness across diverse news domains. Looking ahead, our vision encompasses the expansion of summarization capabilities to encompass news articles in regional languages, such as Tamil, thereby fostering broader inclusivity and global awareness. By embracing linguistic diversity, we aim to facilitate access to vital news information for individuals across varied linguistic backgrounds, contributing to a more informed and connected global community. These future endeavors not only promise to enhance the efficacy of our summarization model but also hold the potential to catalyze advancements in natural language processing and cross-linguistic communication, shaping the future landscape of information dissemination and comprehension.

VII. REFERENCES

[1] Adhika Pramita Widyassari, Supriadi Rustad, "Review of Automatic Text Summarization Techniques", Volume - 34, PP.457 – 480, Feb-2020

[2] A.C.Sanders, R. C. White, L. S. Severson, R. Ma, R. McQueen, H. C. A. Paulo, Y. Zhang, J. S. Erickson, and K. P. Bennett, ''Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse,'' AMIA J. Summits Transl. Sci. Proc., vol. 2021, pp. 555–564, May 2021.

[3] Barbara Probierz, Piotr Stefański, Jan Kozak, "Rapid Detection of Fake News based on Machine Learning", Volume – 56, PP.715-731, Dec-2021

[4] Banerjee, S.; Mitra, P.; Sugiyama, K. "Multi-Document Abstractive Summarization Using ILP Based Multi-Sentence Compression." In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina, 25–31 July 2015; pp. 1208–1214.

[5] Durrett, G.; Berg-Kirkpatrick, T.; Klein, D. "Learning-Based Single-Document Summarization with Compression and Anaphoricity Constraints." arXiv 2016, arXiv:1603.08887

[6] E. Hovy and C.-Y. Lin, ''Automated text summarization and the SUMMARIST system,'' in Proc. Tipster Text Program Phase III, 1996, pp. 197–214, doi: 10.3115/1119089.1119121.

[7] G.K. Kumar and D. M. Rani, ''Paragraph summarization based on word frequency using NLP techniques,'' AIP Conf. Proc., vol. 2317, Feb. 2021, Art. no. 060001, doi: 10.1063/5.0037283.

[8] H.P. Luhn, ''The automatic creation of literature abstracts,'' IBM J. Res. Develop., vol. 2, no. 2, pp. 159–165, Apr. 1958.

[9] Huija Yu, Chang Yue, "News Article Summarization: Attention-based Deep Recurrent Neural Networks", Volume – 3, PP. 145-155, Mar- 2018

[10] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf and Muhammad Ovais Ahmad, "Fake News Detection. Using Machine Learning Ensemble Methods", Volume 54, PP. 378 – 394, Nov -2020

[11] Ji, X.; Zhao, W. "SKGSUM: Abstractive Document Summarization with Semantic Knowledge Graphs." In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021. [CrossRef]

[12] Laxmi B Rananavare, P.Venkata Subba Reddy, "Automatic News Article Summarization", Paper Volume-6, PP. 2347-2693, Jan-2018

[13] Li, W.; Xiao, X.; Liu, J.; Wu, H.; Du, J. "Leveraging Graph to Improve Abstractive Multi-Document Summarization." arXiv 2020, arXiv:2005.10043.

[14] Li, W.; Zhuge, H. "Abstractive Multi-Document Summarization Based on Semantic Link Network." IEEE Trans. Knowl. Data Eng. 2021, 33, 43–54. [CrossRef]

[15] Lene Casey, Emma Davidson, Michael Poon, Hang Dong, "A Review of Natural Language Processing Research", BMC Medical Informatics and Decision Making 21, Volume – 179, PP. 115 - 139 (2019)

[16] Mengyun Cao, Hai Zhuge, "Grouping sentences as better language unit for extractive Text Summarization", Volume – 109, PP.124 – 140, May- 2020

[17] N. Alami, M. Meknassi, N. En-nahnahi, Y. El Adlouni, and O. Ammor, ''Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling,'' Expert Syst. Appl., vol. 172, Jun. 2021, Art. no. 114652, doi: 10.1016/j.eswa.2021.114652.

[18] Ritwik Mishraa, Tirthankar Gayenb, "Automatic lossless-summarization of News Articles", Volume – 62, PP. 630 – 644, July-2021

[19] Sukirti Verma and VagishaNidhi, "Extractive Summarization Using Deep Learning", International Conference On Computational Linguistics and Intelligent Text Processing, 2017.

[20] Sumit Chopra, Michael Auli, "Abstractive Sentence Summarization with Attentive Recurrent Neural Networks", Volume – 7, PP. 995-1012, Oct-2019

[21] Zhu, T.; Hua, W.; Qu, J.; Hosseini, S.; Zhou, X. "Auto-regressive extractive summarization with replacement." World Wide Web 26(4):1- 24, Dec-2022