iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VI **Month of publication:** June 2024

DOI: https://doi.org/10.22214/ijraset.2024.63359

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 VI June 2024- Available at www.ijraset.com

Advancements and Methodologies in Natural Language Processing and Machine Learning: A Comprehensive Review

Kumar Pritam[1], Dr. Abha Chaubey[2], Dr. Avinash Dhole[3]
Computer Science Department, Shri Shankaracharya Technical Campus, Bhilai

Abstract: ***This review explores notable advancements and methodologies in Natural Language Processing (NLP) and machine learning. Key innovations include convolutional neural networks capable of generating multiple language processing predictions and the Transformer-XL model, which enhances temporal coherence and dependency handling. The importance of logical reasoning and common sense in AI is emphasized, along with practical methods like web mining for building extensive knowledge repositories. The development of NLP tools for specific languages, such as Gujarati, and the utilization of BERT for bidirectional training are discussed, showcasing significant performance improvements across multiple tasks. Sentiment analysis is highlighted as a crucial tool for interpreting unstructured data from social media, achieving high accuracy through rule-based approaches and feature extraction methods. The role of social media in disseminating emotionally charged content and its impact on user behavior is examined, providing valuable insights for businesses to enhance marketing strategies. Additionally, the review addresses the challenge of early risk detection on social media platforms, emphasizing the need for effective pre-processing techniques to handle diverse and unstructured content. The integration of deep learning techniques in sentiment analysis is shown to offer competitive advantages, revealing consumer preferences and product trends.***

Keywords: ***NLP, Text summarization***

I. INTRODUCTION

The multidisciplinary field of natural language processing (NLP) seeks to enable robotics with the ability to understand, interpret, and generate meaningful human language that is relevant to the given context. The field of study combines concepts from cognitive psychology, computer science, linguistics, and artificial intelligence in order to establish a correlation between human communication and machine comprehension. Natural language processing (NLP) technologies and approaches enable computers to comprehend and analyze unstructured text input, facilitating a wide range of applications that enhance human-computer interaction. The significance of natural language processing (NLP) lies in its capacity to transform human-machine interaction by enabling computers to generate and comprehend human language at a level that was previously regarded as science fiction. The integration of Natural Language Processing (NLP) into a variety of applications across various industries has significantly transformed the effectiveness and efficiency of various tasks. Natural language processing (NLP) has exhibited its efficacy in automating procedures, extracting information from textual data, and establishing novel means of communication. There exist multiple applications of this technology, including the utilization of chatbots to improve customer service and the implementation of language-based analysis to aid in medical diagnosis. Natural language processing (NLP) has a significant impact on various sectors such as finance, healthcare, e-commerce, education, and entertainment. Sentiment analysis aids organizations in evaluating public sentiment, while machine translation eradicates linguistic barriers across nations. Natural Language Processing (NLP) is driving innovation and transforming organizations in various domains such as text summarization, voice recognition, and language generation.

II. LITERATURE REVIEW

Alshemali et. al. [1] The deep learning models have demonstrated efficacy in addressing various challenges in the field of natural language processing (NLP). However, an increasing amount of scholarly research suggests that hostile entities can easily manipulate deep neural networks (DNNs). Adversarial examples are inputs that have been intentionally modified to deceive a target model into producing deceptive results. These inputs undergo modest alterations in order to achieve their deceptive nature. The utilization of neural networks in safety-critical environments is greatly hindered by their susceptibility to adversarial examples. This paper presents a thorough examination of the use of adversarial scenarios in the domain of natural language processing (NLP) for the purpose of improving the resilience of deep neural networks (DNNs).

©IJRASET: All Rights are Reserved

The study also examines the existing techniques utilized to alleviate the impacts of deep neural networks. This article provides a detailed examination of the existing methods used to create adversarial texts and presents a systematic approach for categorizing them. In addition, we provide a thorough analysis of the obstacles that different defensive strategies face when dealing with adversarial events and propose pertinent areas for future investigation.

Baclic et. al. [2] The construction and comprehension of human language are the main goals of natural language processing (NLP), which is a subfield of artificial intelligence. The capability to effectively analyze large volumes of text has been significantly enhanced by advancements in natural language processing (NLP). The technological progress has facilitated novel prospects for conducting health-related investigations and formulating evidence-based decisions. Scientific journals, technical reports, health records, social media, surveys, registries, and other documents are considered primary sources for acquiring valuable public health data. One of the proposed enhancements is the optimization of the existing monitoring systems to enable the swift identification of susceptible individuals and ailments. Furthermore, there are continuous efforts in research and development focused on enhancing disease prevention techniques. This involves the assessment of the therapy's effectiveness and safety. The goal is to enhance the dissemination of health information by delivering dependable answers to inquiries pertaining to health. Natural Language Processing (NLP) is gaining recognition among public health professionals as a valuable method for tackling the pervasive occurrence of health inequalities and inequities in the population. The purpose of this essay is to elucidate the potential benefits and drawbacks of integrating natural language processing into the domain of public health.

Bahdanau et. al. [3] As an alternative to traditional machine translation methods, a novel technique called neural machine translation has been suggested. Contrary to statistical machine translation, the goal of neural machine translation is to develop a single neural network that can be collectively adjusted to achieve optimal translation outcomes. Recently, encoder-decoders have emerged as a dominant classification for neural machine translation models. An encoder is a hardware or software component that transforms an input sequence or phrase into a fixed-length vector representation. The vector is subsequently utilized by the decoder to produce a translation. The aim of this research is to improve the underlying encoder-decoder architecture by allowing a model to independently choose important segments of a source phrase to forecast a target word. As a result, there will be no need to explicitly classify these components as separate entities. The requirement to use a vector with a pre-defined length is probably the cause of the difficulty in enhancing the effectiveness of this framework. The current methodology facilitates the effective transformation of English texts into French, attaining a performance level that is on par with the most advanced phrase-based system. Furthermore, a comprehensive analysis demonstrates that the model's hard and soft alignments are closely aligned with our intuitive comprehension.

Chalkidis et. al. [4] The BERT (Bidirectional Encoder Representations from Transformers) model has exhibited remarkable effectiveness in a range of natural language processing (NLP) tasks. There is a lack of extensive research that examines the principles of adaptation in relation to this specific phenomenon in specific regions. The investigation is primarily focused on the legal domain. This study examines different methodologies for evaluating diverse datasets using BERT models, with the objective of being sufficiently prepared for potential legal disputes that may occur in the future. The findings of our investigation suggest that the general acceptance of commonly used pre-training and fine-tuning techniques does not consistently produce positive results in the legal industry. It is recommended to perform a thorough analysis of the different strategies that can be utilized to integrate BERT into specific domains. The available options at present are as follows: The potential alternatives are as follows: the first option is to retain BERT in its current state, the second option is to enhance BERT by integrating supplementary pre-training with domain-specific datasets, and the third option is to commence BERT's pre-training from scratch using domain-specific datasets. Moreover, we provide LEGAL-BERT, a compilation of BERT models that have been purposefully developed to expedite the progress of legal NLP research, computational law, and legal technology applications. In addition, it is advisable to broaden the hyper-parameter search options in order to increase the potential for more effective future research.

Cho et. al. [5] Neural machine translation is an advanced computational technique that utilizes neural networks exclusively for the purpose of statistical analysis. Neural machine translation models usually consist of two essential components: an encoder and a decoder. The encoder utilizes a computational method to generate a standardized representation of fixed length. This is achieved by supplying the encoder with input text of different durations. The decoder utilizes this representation to ensure an accurate translation. The objective of this investigation is to assess the translation capabilities of neural machine systems. In order to accomplish this, we will evaluate the effectiveness of two separate models: an RNN encoder-decoder and a recently developed gated recursive convolutional neural network. Based on our research findings, neural machine translation demonstrates a high level of effectiveness when it comes to translating concise phrases that have limited or no elements of a foreign language.

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

However, the efficacy of sentences is significantly diminished as their length increases and they include a greater quantity of foreign words. Furthermore, it is apparent that the syntactic organization of the text can be readily comprehended by the suggested gated recursive convolutional network.

Chouikhi et. al. [6] The accuracy of sentiment analysis techniques has been the focus of extensive research, involving various implementations, ranging from simple linear models to intricate deep neural network implementations. The transformer-based paradigm is widely recognized as the state-of-the-art approach for facilitating multilingual support, encompassing languages such as Arabic, Turkish, German, French, and English. The sentiment analysis performance of the system has exhibited a satisfactory level of accuracy and effectiveness. The accuracy of Arabic sentiment analysis can be enhanced by increasing the level of tokenization utilized during the data processing phase. Users may encounter challenges due to the intricate nature of Arabic's grammatical structure, the abundance of dialects, and the restricted accessibility of information. The original BERT tokenizer has been substituted with the Arabic BERT tokenizer, resulting in enhancements to the suggested solution. An extensive range of cases, encompassing both traditional and geographical variations, were analyzed using various methodologies. The results suggest that the random search method was effectively combined with hyperparameter optimization to achieve optimal outcomes across multiple datasets. Regarding both classification accuracy and quality, the experimental evaluation results demonstrate that the proposed technique outperforms the Arabic BERT and AraBERT models.

Collobert et. al. [7] The proposed architecture exhibits the capability to generate diverse language processing predictions based on a given word. To enhance the semantic and grammatical understanding of a sentence, a language model generates predictions that include various identifiers and information. The multimodal learning model involves training the entire network on multiple tasks, which is referred to as the weight-sharing technique. For all tasks, except the language model, labeled data is a necessary requirement. The language model utilizes unannotated text input to provide a distinctive approach to semi-supervised learning in collaborative tasks. The following text provides an example of an exceptional achievement that was facilitated by the utilization of multitasking techniques and semi-supervised learning. The objective of this achievement was to improve the ability to adapt collective responsibilities to a range of tasks.

Dai et. al. [8] Transformers possess the capability to foster long-lasting relationships. However, there is a predefined temporal constraint on the amount of language modeling that can be obtained. The main emphasis of this concept is on the brain architecture of Transformer-XL, aiming to enhance the duration of information retention and dependency, as well as improve temporal coherence. The system employs a unique algorithm for encoding positions and a methodology for recurring at the segment level. The methodology employed in this study is specifically designed to efficiently tackle the problem of fragmented context, while also considering the extended reliance on a particular approach. The Transformer-XL model surpasses traditional Transformers in effectively handling sequences with different lengths, showcasing better performance in detecting dependencies compared to Recurrent Neural Networks (RNNs), and also outperforming them in terms of time efficiency. Furthermore, it achieves assessment rates that are 1,800 times faster than those of traditional transformers. The model demonstrates outstanding performance on multiple benchmark datasets, including enwiki8, text8, WikiText-103, and One Billion Words. The model achieves a score of 21.8 on the Penn Treebank dataset and 0.99 on the enwiki8 dataset without any modifications or enhancements. The Transformer-XL model demonstrates its ability to generate text articles that are coherent and cohesive, even with a large number of tokens. The machine's performance is apparent even when it is trained exclusively with the WikiText-103 dataset. Access to the source code, hyperparameters, and pre-trained models within the Tensorflow and PyTorch frameworks is provided to users via our platform.

Marcus et. al. [9] performed a study to evaluate the significance of common sense and reasoning in the field of artificial intelligence. Proficiency in practical knowledge and cognitive abilities is essential for tasks involving strategic planning, comprehension of textual information, application of logical reasoning, and analysis of visual data. Web mining, crowdsourcing, logical analysis, and manual construction of large-scale information libraries are all viable techniques for collecting and organizing large amounts of data. Notwithstanding the fact that intelligent algorithms may carry out these tasks without exactly replicating human intelligence, it is crucial to possess a comprehensive understanding of human cognition.

Dabhi et. al. [10] Natural Language Processing (NLP) is a branch of artificial intelligence that involves the use of computational algorithms to analyze and comprehend human language. Software is utilized in this process. Various tasks can be automated by leveraging natural language processing (NLP) technologies. These occupations require a diverse range of technical skills. The capabilities enumerated in this list encompass language translation, data extraction and summarization from complex databases, spam email screening, false information detection, political sentiment analysis, policy analysis, and leveraging patient histories for delivering high-quality healthcare. Gujarati, an Indo-Aryan language, is spoken by over sixty million individuals globally.

There are ongoing significant efforts to develop Natural Language Processing (NLP) tools and software that are specifically designed for Indian languages. The study thoroughly analyzes and categorizes the materials and processes used in the development of Natural Language Processing (NLP) system components for the Gujarati language. Further investigation was conducted following the assessment of multiple widely used open-source software applications. Potential strategies for mitigating the challenges related to the development of components and materials for the Gujarati Natural Language Processing (NLP) system will be incorporated into the supplementary topics for discussion. This study provides valuable information for academicians, professionals, and researchers who are interested in gaining a comprehensive understanding of the potential, challenges, and technological requirements involved in developing Gujarati natural language processing (NLP) systems.

Devlin et. al. [11] The BERT model was specifically designed for the purpose of pre-training deep bidirectional representations from untagged text. Unlike previous language representation models, this model undergoes simultaneous training on both the left and right context at every level. The BERT model can be enhanced by adding an extra output layer, which can lead to improvements without requiring major changes to the task-specific structure. This enhancement enables the creation of advanced models that can be utilized for various applications, such as language inference and query answering. From both theoretical and practical perspectives, the BERT model demonstrates robustness and reliability. The system exhibits outstanding performance on eleven cutting-edge tasks in the field of natural language processing. The model significantly improves the accuracy of the SQuAD v1.1 question answering Test F1 metric to 93.2, the GLUE score to 80.5%, the MultiNLI to 86.7%, and the SQuAD v2.0 Test F1 metric to 83.1. This enhancement results in a 5.1 percentage point increment.

Anjaria et. al. [12] Sentiment analysis is a critical aspect of analyzing unstructured data generated by social networking platforms. The sentiment analysis process categorizes textual phrases into various sentiment polarities, such as positive, negative, neutral, jubilant, enraged, repulsed, fearful, and others. The study employed a text summarization technique to analyze the polarity (positive, negative, or neutral) of emotions expressed in Bangla texts. Utilizing a rule-based methodology and manually designed characteristics, the Bangla dataset attained a remarkable accuracy rate of 98.33%. The study demonstrated a notable maximum accuracy of 98.33% in accurately classifying Bangla blog entries based on their emotional polarity. This was achieved by employing a feature extraction technique that heavily relied on text summarization. This study assesses the sentiment in Bangla writings using a rule-based algorithm and a combination of attributes generated by humans. This initiative will provide researchers from diverse fields with the opportunity to enhance their understanding and gain fresh insights. Consequently, researchers will be afforded the opportunity to perform a more comprehensive investigation of topics related to sentiment analysis.

Ramaswamy et. al. [13] It is crucial for any organization to prioritize consumer satisfaction and have a comprehensive understanding of consumer behavior in order to maintain a competitive advantage in the market. In order to enhance the accuracy and efficiency of product development and marketing strategies, organizations need to possess a comprehensive understanding of the prevailing consumer sentiments. There are various techniques available for acquiring customer assessments. Our main focus is on unstructured data, particularly textual content that is obtained from various sources, such as social media, survey responses, audio recordings of customer interactions, and text transcripts of conversations. An accurate assessment of this data is crucial, as it can provide organizations with a significant competitive edge and reveal insights on various subjects, including consumer behavior and product shortcomings. The organization would gain significant benefits by enhancing its data collection capabilities pertaining to consumer preferences, product updates, and marketing insights. The objective of this study is to explore different Deep Learning and Natural Language Processing (NLP) methodologies to enhance the assessment of contextual data and the acquisition of customer feedback.

Stieglitz et. al. [14] Social media, a novel communication technique, has enabled the transmission of information via social networks. Several variables, including consumer behavior, network characteristics, and content, have been identified in previous studies as potential factors that can impact the spread of information. The correlation between emotional states and the transmission of information within the context of social media has been largely overlooked in academic studies. The objective of this study is to examine potential associations between the emotional content expressed in online media and the sharing behavior of users on social media platforms. The research primarily focuses on the political communication sphere of Twitter. The analysis of two datasets, consisting of over 165,000 tweets, reveals that emotionally charged comments on Twitter have a higher likelihood of being retweeted at a faster pace compared to neutral ones. It is recommended that companies prioritize the development of advertising content that elicits powerful emotional responses. Enterprises should give priority to assessing the sentiment surrounding their brands and products through social media communication.

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

Ramasamy et. al. [15] Businesses and organizations have consistently showcased the significance and impact of the community's opinions and contributions. Organizations now have an enhanced capability to analyze and assess a wide range of factors that were previously examined using time-consuming, non-traditional, and unreliable techniques, due to the widespread adoption of social media. The concept of "sentiment analysis" is intricately connected to this analytical methodology. Sentiment analysis is a multidisciplinary field that focuses on the effective categorization of user-generated content into pre-defined polarities. Sentiment analysis and detection can be implemented using various methodologies and tools. One method involves the application of supervised machine learning algorithms that are trained on a specific set of training data from the target corpus. The algorithms possess the ability to effectively classify novel data by acquiring the capability to recognize patterns and extract relevant information through the training process. The process of classifying data using annotated corpora that depend on dictionaries is the fundamental element of lexical techniques. On the other hand, hybrid tools integrate lexicon-based algorithms with machine learning techniques. The Weka software is utilized in this study for the purpose of performing sentiment analysis, utilizing Support Vector Machines (SVM) as the underlying algorithm. In the domain of supervised machine learning, support vector machines (SVMs) are commonly utilized for the purpose of ascertaining the polarity of text. Support Vector Machines (SVM) are evaluated using two preclassified datasets acquired from Twitter. Three metrics, namely F-Measure, Precision, and Recall, are utilized for the purpose of comparison. The findings are visually represented through the use of tables and graphs.

Burdisso et. al. [16] The increasing prevalence of the Internet has highlighted the need for the creation of intelligent systems that can effectively tackle early risk detection (ERD) challenges on social media platforms. These challenges include the timely identification of depression, the early detection of false information, and the recognition of sexual predators. Addressing these challenges has become increasingly crucial. Contemporary systems heavily rely on machine learning techniques and need to effectively handle ongoing data streams, as users continuously contribute their data. In addition, it is necessary for these systems to have the capability to determine the threshold at which a sufficient amount of processed data is present in order to accurately classify individuals. In addition, it is crucial for these systems to provide evidence to support their decisions. This is because ERD responsibilities involve making decisions that have the potential to be hazardous and can affect the lives of individuals. However, most supervised machine learning models, regardless of whether they are modern or traditional, are insufficient for effectively handling this particular scenario. The observed outcome can be attributed to the inherent tendency of these systems to operate in an opaque manner or their limited capability to facilitate progressive classification or learning. The SS3 model is an exceptional supervised learning model that is specifically designed to accommodate various aspects of text classification. The main objective of SS3 is to provide a comprehensive framework for efficiently addressing concerns related to Entity-Relationship Diagram (ERD). The model was evaluated by utilizing the eRisk2017 pilot challenge from the Conference and Labs of the Evaluation Forum (CLEF). This challenge is specifically focused on the early diagnosis of depression. The majority of the thirty entries submitted for this competition consisted of contemporary methodologies. The classifier we developed demonstrates superior performance compared to existing models, as supported by empirical evidence. It is designed to be less computationally intensive than standard classifiers and has the capability to provide explanations for its reasoning.

Palomino et. al. [17] The field of sentiment analysis is experiencing rapid expansion, fueled by practical and academic needs. Preprocessing techniques play a crucial role in data cleansing, especially when conducting sentiment analysis on social media platforms. This is because the content on these platforms often ignores grammatical and orthographic rules. Prior to commencing the analysis, it is crucial to carry out pre-processing in order to standardize the text. This is necessary because social media platforms often contain a multitude of abbreviations, shortened words, emoticons, avatars, and specialized language. Pre-processing has been extensively studied in academic literature and is widely recognized as a crucial step. However, there is a lack of consensus regarding the most effective methods. In order to obtain accurate results, a quantitative analysis of various combinations of pre-processing components was conducted. Additionally, a thorough review of the existing literature on the subject was performed. The focus of our attention has been directed towards sentiment analysis on this website. This is because Twitter is widely recognized as a prominent source of publicly accessible information. Additionally, we conducted an assessment of the influence of numerous commercially available solutions and our proprietary algorithm on the overall accuracy of different combinations of pre-processing components. The investigation findings confirm that the organization of pre-processing components is crucial and significantly improves the effectiveness of naïve Bayes classifiers. Additionally, it has been verified that the process of lemmatization enhances the efficiency of indexing. However, it has been observed that it does not have a significant impact on the quality of sentiment analysis.

Haddi et. al. [18] The comprehension of current trends and the general attitude towards things can be challenging due to the vast diversity and volume of social media data. Therefore, there is a need for the automated and real-time extraction and mining of opinions. Online opinion mining is considered a complex text classification challenge within the field of sentiment analysis. This study examines the role of text pre-processing in sentiment analysis and provides experimental findings that demonstrate the potential for significantly improving sentiment analysis using support vector machines (SVM) by employing appropriate feature selection and representation techniques. Sentiment analysis is often considered a challenging problem in the literature. However, it has been shown to achieve a similar level of accuracy as topic categorization.

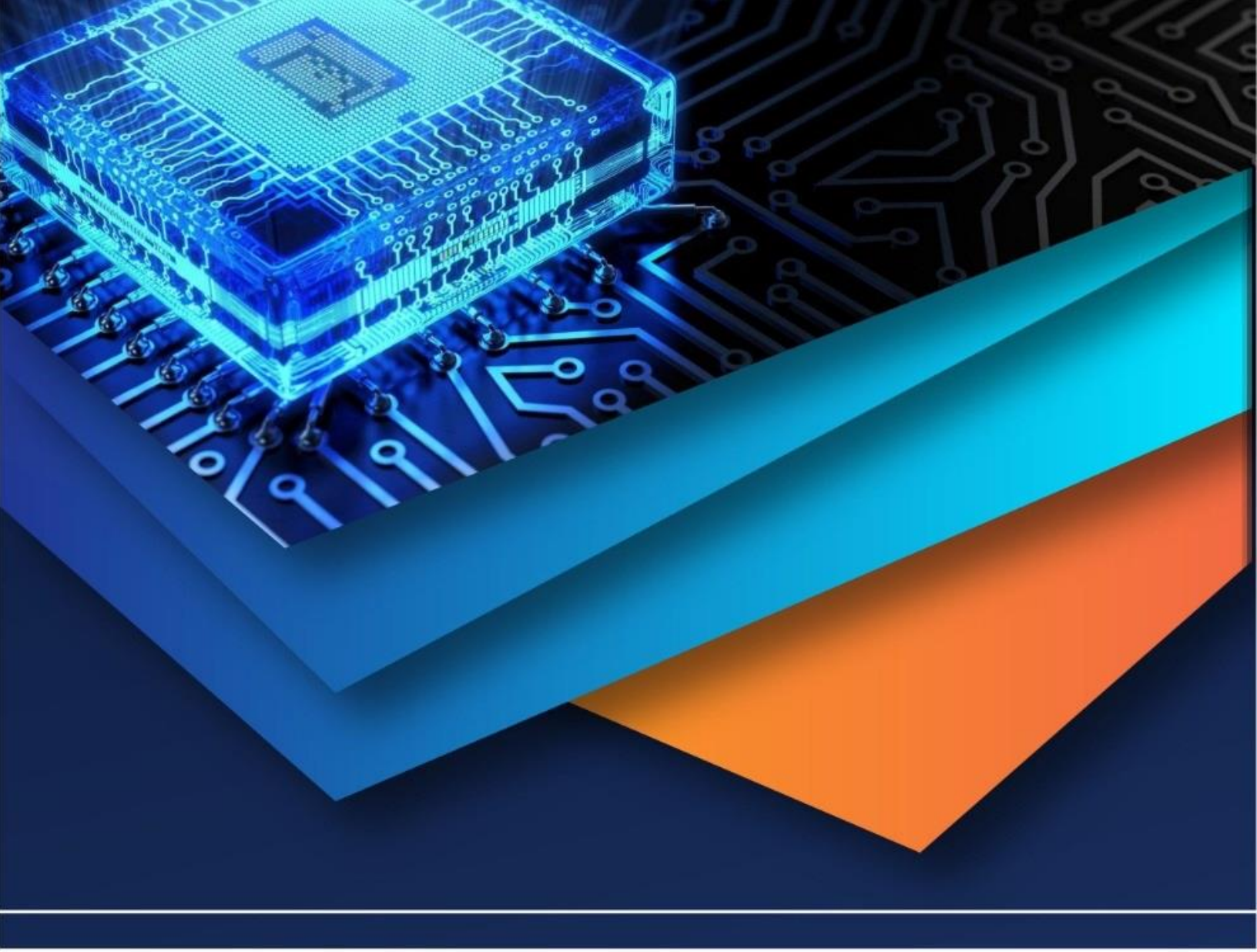
III.CONCLUSION

The reviewed papers highlight significant advancements and methodologies in the realm of Natural Language Processing (NLP) and machine learning. Notably, convolutional neural networks have been shown to generate multiple language processing predictions, enhancing semantic and grammatical understanding through a multimodal learning strategy. Innovations like Transformer-XL demonstrate superior temporal coherence and dependency handling, significantly outperforming traditional transformers and RNNs in various tasks. The importance of logical reasoning and common sense in AI has been emphasized, alongside the role of practical methods like web mining and crowdsourcing for building extensive knowledge repositories. The field of NLP extends to a variety of applications, including the development of tools tailored for specific languages like Gujarati, and leveraging BERT's bidirectional training to achieve state-of-the-art performance across multiple tasks. Sentiment analysis emerges as a critical tool for decoding unstructured data from social media, with impressive accuracy rates achieved using rule-based approaches and feature extraction methods. The integration of deep learning techniques is vital for assessing customer feedback and improving product development strategies.

REFERENCES

[1] Alshemali B, Kalita J (2020) Improving the reliability of deep neural networks in NLP: A review. KnowlBased Syst 191:105
[2] Baclic O, Tunis M, Young K, Doan C, Swerdfeger H, Schonfeld J (2020) Artificial intelligence in public health: challenges and opportunities for public health made possible by advances in natural language processing. Can Commun Dis Rep 46(6):161
[3] Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In ICLR 2015
[4] Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I (2020) LEGAL-BERT: the muppets straight out of law school. arXiv preprint arXiv:2010.02559
[5] Cho K, Van Merriënboer B, Bahdanau D, Bengio Y, (2014) On the properties of neural machine translation: encoder-decoder approaches. arXiv preprint arXiv:1409.1259
[6] Chouikhi H, Chniter H, Jarray F (2021) Arabic sentiment analysis using BERT model. In international conference on computational collective intelligence (pp. 621-632). Springer, Cham
[7] Collobert R, Weston J (2008) A unified architecture for natural language processing. In proceedings of the 25th international conference on machine learning (pp. 160–167)
[8] Dai Z, Yang Z, Yang Y, Carbonell J, Le QV, Salakhutdinov R, (2019) Transformer-xl: attentive language models beyond a fixed-length context. arXiv preprint arXiv:1901.02860
[9] Davis E, Marcus G (2015) Commonsense reasoning and commonsense knowledge in artificial intelligence. Commun ACM 58(9):92–103
[10] Desai NP, Dabhi VK (2022) Resources and components for Gujarati NLP systems: a survey. Artif Intell Rev:1–19
[11] Devlin J, Chang MW, Lee K, Toutanova K, (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
[12] Anjaria, M.; Guddeti, R.M.R. A novel sentiment analysis of social networks using supervised learning. Soc. Netw. Anal. Min. 2014, 4, 181. [CrossRef]
[13] Ramaswamy, S.; DeClerck, N. Customer perception analysis using deep learning and NLP. Procedia Comput. Sci. 2018, 140, 170–178. [CrossRef]
[14] Stieglitz, S.; Dang-Xuan, L. Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. J. Manag. Inf. Syst. 2013, 29, 217–248
[15] Ramasamy, L.K.; Kadry, S.; Nam, Y.; Meqdad, M.N. Performance analysis of sentiments in Twitter dataset using SVM models. Int. J. Electr. Comput. Eng. 2021, 11, 2275–2284.
[16] Burdisso, S.G.; Errecalde, M.; Montes-y-Gómez, M. A text classification framework for simple and effective early depression detection over social media streams. Expert Syst. Appl. 2019, 133, 182–197.
[17] Palomino, M.A.; Aider, F. Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis. Appl. Sci. 2022, 12, 8765.
[18] Haddi, E.; Liu, X.; Shi, Y. The role of text pre-processing in sentiment analysis. Procedia Comput. Sci. 2013, 17, 26–32

©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)