IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

GENERATION OF TITLE USING SENTIMENTAL ANALYSIS

Dr.M.Priyadharshini, Sirpy Palaniswamy, Sanjit N M, Suriya J

Assistant Professor, UG Scholar, UG Scholar, UG Scholar
Information Technology,
Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu

Abstract: In the digital era, the vast array of online content poses a significant challenge for users seeking relevant information, primarily due to the absence of emotionally resonant titles. Existing title generation methods falter in capturing nuanced emotional tones, resulting in diminished user engagement. Moreover, current systems lack adaptability to user preferences and real-time trends, thereby limiting their efficacy across diverse audiences. This paper advocates for a novel title generation system that harnesses advanced sentiment analysis techniques and innovative language generation strategies to craft captivating, emotionally resonant titles. The proposed system aims to accurately reflect content sentiment. Furthermore, it seeks to dynamically adapt to evolving content landscapes, thereby enhancing content discoverability and enriching the online browsing experience across various platforms.

***Index Terms* -** Sentiment Analysis, NLP, Title Generation

I. INTRODUCTION

In today's digital age, where information overload is the norm, the importance of crafting compelling and engaging titles for online content cannot be overstated. A well-crafted title serves as the gateway to your content, capturing the attention of readers and enticing them to explore further. However, generating such titles is not a trivial task, particularly in the absence of a structured methodology. This paper addresses this challenge by proposing a novel approach to title generation leveraging the power of sentiment analysis.

The significance of sentiment analysis in understanding user preferences and emotions cannot be overlooked. By analyzing the sentiment expressed in textual content, sentiment analysis algorithms can uncover valuable insights into audience preferences, helping content creators tailor their titles to resonate more deeply with their target audience. This paper explores the potential of sentiment analysis as a tool for generating titles that not only capture attention but also evoke emotion and drive engagement.

At the heart of our proposed approach lies the recognition that titles are more than just strings of words – they are powerful tools for conveying emotion, capturing attention, and shaping reader perceptions. By leveraging sentiment analysis techniques, we aim to imbue titles with emotional depth, allowing them to resonate more deeply with readers and elicit stronger emotional responses.

Furthermore, our approach emphasizes the importance of personalization in title generation. Recognizing that different readers may respond differently to the same title, we propose a methodology for dynamically adapting titles based on individual reader preferences and emotional states. By tailoring titles to the unique preferences and emotions of each reader, we aim to create a more personalized and engaging browsing experience.

In addition to personalization, our approach also considers the dynamic nature of online content consumption trends. By monitoring real-time sentiment trends and incorporating them into title generation algorithms, we aim to ensure that titles remain timely, relevant, and reflective of current emotional states and cultural contexts.

Overall, this paper presents a comprehensive framework for the generation of titles using sentiment analysis. By harnessing the power of sentiment analysis, we aim to revolutionize the way titles are generated, creating titles that not only capture attention but also evoke emotion, drive engagement, and resonate deeply with readers. We used Bi-Directional LSTM for sentiment analysis and transformers to develop the framework.

II.RELATED WORKS

Generation of titles for a blog or an article has been in the works for a very long time now, and looking at the various methods for doing so is very interesting and challenging at the same time. Title generation can be achieved through methods like: transformers based summarization, which summarizes a given text to give a very short abstract of the article and thereby generating a short title, Generative Adversarial Network (GAN), transformers based Seq2Seq, and so on. In this section, we present some of the previous studies and works in title generation and sentiment analysis

A. Transformers Based Title Generation

In recent years, the field of natural language processing (NLP) has witnessed remarkable advancements, particularly with the advent of Transformer-based models. Transformers, introduced by Vaswani et al. in the seminal paper "Attention is All You Need," [1] have revolutionized various NLP tasks, including language translation, text summarization, and sentiment analysis. Among these tasks, title generation stands out as a crucial aspect of content creation, particularly in the context of digital media, academic research, and marketing.

Transformers, such as the Bidirectional Encoder Representations from Transformers (BERT)[2] and the Generative Pre-trained Transformer (GPT) [3], have emerged as powerful tools for title generation. These models leverage self-attention mechanisms to capture contextual dependencies within text data, enabling them to generate titles that are not only grammatically correct but also semantically coherent and contextually relevant.

One of the key advantages of using Transformers for title generation is their ability to capture long-range dependencies and semantic relationships within text data [1]. Unlike traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs), which may struggle with capturing such dependencies over long sequences, Transformers excel at processing large amounts of text data efficiently.

Moreover, Transformers can be fine-tuned on domain-specific datasets to generate titles that adhere to specific style guidelines, terminology, and audience preferences [4]. This adaptability makes them suitable for a wide range of applications, including news article headlines, blog post titles, academic paper titles, and advertising copy.

Transformers have demonstrated their efficacy in various NLP tasks, and their application to title generation holds promise for improving content discoverability, user engagement, and overall readability in diverse digital media contexts.

B. Title Generation

Several title creation methods have been investigated in the past. Alternative methods, including kNN [5], have been proposed instead of neural networks. There has been study on title generating using Transformer [6-8].

Previous study on title generating has primarily focused on news. The headline generation approach is not suitable for generating titles for journal papers due to their tendency to be over-expressed for mass consumption.

To create a title for a journal paper, consider creating numerous options and selecting one [9]. Although this strategy generates candidates, it may not always identify relevant content in the abstract due to the length of journal articles.

C. Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a computational technique employed to discern the sentiment or emotional tone expressed within a piece of text [10]. Its primary objective is to classify text data into positive, negative, or neutral categories based on the emotions conveyed. This analytical approach has garnered substantial significance across various domains, including social media monitoring, customer feedback analysis, market research, and product reviews.

In the realm of sentiment analysis, numerous methodologies and algorithms have been developed to address the challenges associated with accurately deciphering human emotions from natural language text. Early approaches predominantly relied on rule-based methods [11], where predefined rules and lexicons were utilized to identify sentiment-bearing words and phrases. Concurrently, machine learning and deep learning techniques have gained prominence, leveraging labeled training data to learn intricate patterns and relationships in text data [12].

Traditional machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Logistic Regression have been extensively employed for sentiment analysis tasks[13], relying on handcrafted features extracted from text to classify sentiment. However, the advent of deep learning has revolutionized sentiment analysis, with models like recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformers demonstrating superior performance in capturing complex relationships and semantics within text data [14].

Large-scale labeled datasets, including IMDb movie reviews, Amazon product reviews, and Twitter sentiment datasets, have played a pivotal role in advancing sentiment analysis research, providing researchers with ample resources for training and evaluating models across diverse domains and languages.

We developed a sentiment analysis model based on Bi-directional Long Short-Term Memory (LSTM) networks. These networks have demonstrated effectiveness in capturing long-range dependencies in sequential data, making them suitable for sentiment analysis tasks involving text data with varying lengths and contexts.

III.METHODOLOGY

Our study aims to develop an AI model that can generate a title based on the emotion and the content produced before it. The generic title generation systems out there summarize a given text and generate it as a title given the summary is short enough for a title and fit for the content. However, this poses a problem for the readers to grasp the emotion that is being portrayed in the article, thus a sentiment based title needs to be developed. This section speaks about each model.

A. Dataset

For the developing of a sentiment analysis model, we decided to look into kaggle for the dataset to train our model on. For developing a wider range of emotions, rather than just positive, neutral and negative, we decided to go ahead with a more specific emotions like sadness, joy, anger, etc., for the dataset. The dataset[15] consists we decided on consists of 6 different emotions as shown in Fig. 1.

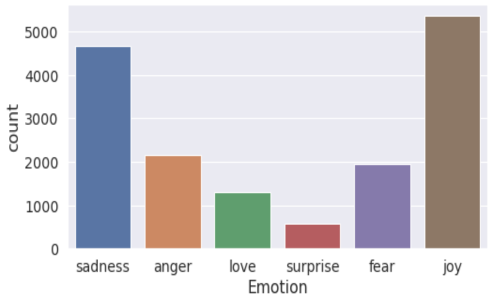

Fig. 1. The instances of emotions in the dataset along with their frequency in the dataset.

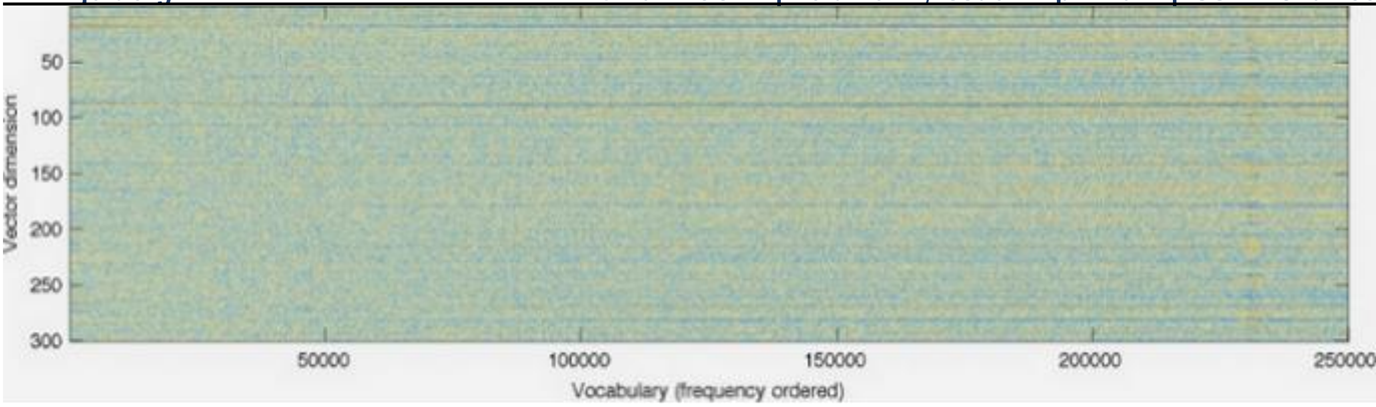

Fig. 2. GloVe produces word vectors with a marked banded structure that is evident upon visualization.

For representation of the textual data, we decided to go with the GloVe dataset for 200 dimensional word embeddings . These word representation from the GloVe dataset, which have been pre-trained on the combined Wikipedia 2014 + Gigaword 5th Edition corpora (6B tokens, 400K vocab) helps us in encoding our words in the article as dense vectors, thus enabling to procure a sentiment for the article. The word vector bands are shown in Fig. 2

B. Model Architecture

For the sentiment analysis model, the Recurrent Neural Network (RNN) based Bidirectional LSTM is the algorithm that we decided would be best using a trial and error method with a BERT model, which ended up giving us a lower accuracy than the Bidirectional LSTM algorithm. The model comprises of an embedding, bidirectional and dense layers.

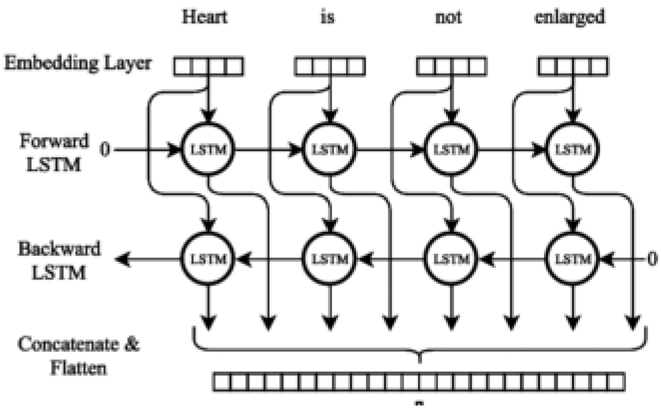

Fig. 3. Bidirectional LSTM algorithm structure

In the case for the title generation model, the concept of pre-trained t5 model was used. The pre-trained t5 model is based on text summarization and was trained on a dataset of medium articles and their respective titles. The selection of a pre-trained t5 model is due to the performance being better in the case of a transformer than a GAN or Seq2Seq technique.

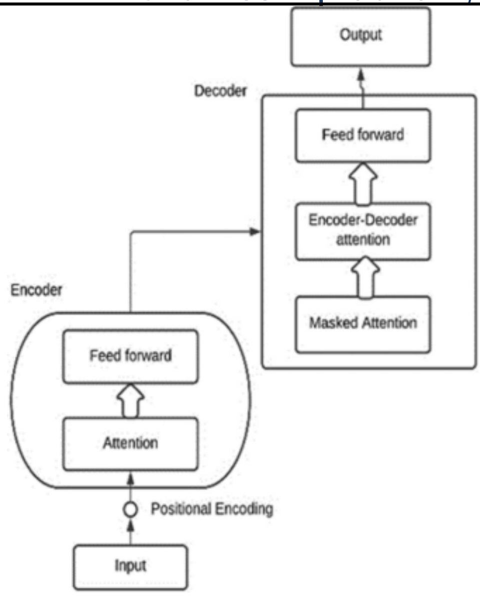

Fig. 4. T5 summarization algorithm structure

C. Training Strategy

For the training of the sentiment analysis model, since the classification of emotion is probability based, the softmax function is the best option for activation function. Hence, we decided to go with the following softmax activation function.

$$\sigma(\vec{Z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

For the training of the sentiment model, we are going with the classic 75-25 split of the training data [15]. With an early stopping callback set, the training of the sentiment analysis model is set to be trained. With the accuracy metrics being monitored, the early stopping callback will prevent from over fitting the data and ruining the model. With the Adam optimizer set on a 0.005 learning rate, the model will be trained on the 75% split of the training data [15].

D. Incorporation of Sentiment in Title Generation

After developing the model for sentiment analysis and also procuring the title generation model, we are proposing to incorporate both of them into one by a simple concatenation code. With the sentiment being procured from the context, the derived sentiment is then sent to the title generator along with the context, which then encodes both inputs and gives out a single title based on the sentiment and the context of the article.

IV.EXPERIMENTS

A. Evaluating the Sentiment Analysis Model

The trained sentiment analysis model is evaluated on the test data and the accuracy was at 93% as shown in Fig. 5.

	precision	recall	f1-score	support
0	0.92	0.95	0.93	275
1	0.89	0.92	0.90	224
2	0.98	0.93	0.95	695
3	0.80	0.94	0.87	159
4	0.96	0.97	0.97	581
5	0.85	0.68	0.76	66
accuracy			0.93	2000
macro avg	0.90	0.90	0.90	2000
weighted avg	0.94	0.93	0.93	2000

Fig. 5. Accuracy, Recall, F1-Score and Support of the sentiment analysis model.

The confusion matrix for the same was mapped along with the labels marked in the Fig. 8. For the more accurate representation of the accuracy, a graph representing the loss and accuracy measure of training versus the validation can be seen in Fig. 6. and Fig. 7.

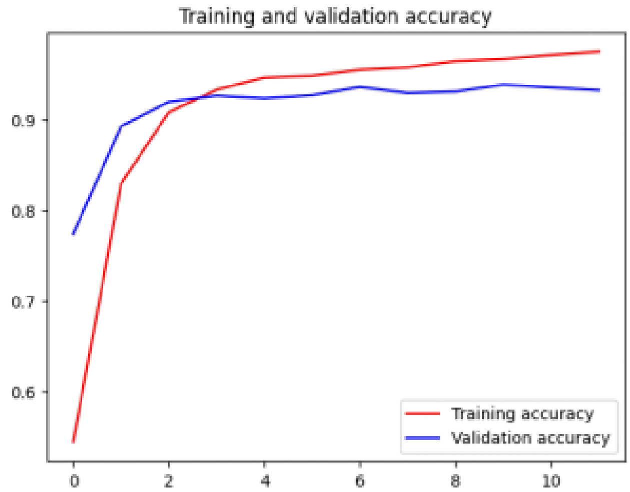

Fig. 6. Training versus validation accuracy graph

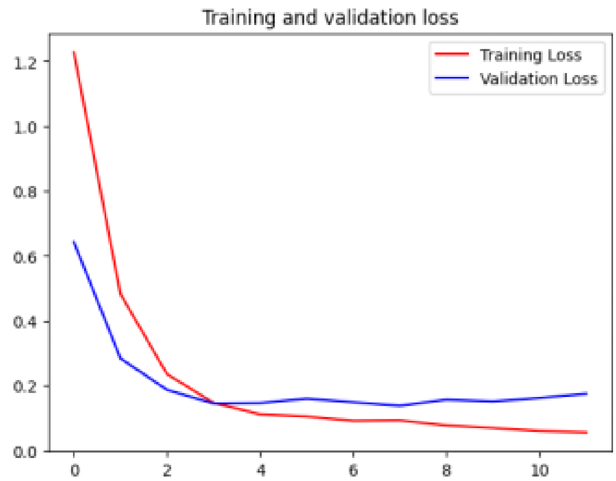

Fig. 7. Training versus validation loss graph

The sentiment analysis model that was developed performed way better than the first model using BERT which procured us only 53% accuracy, so we decided to go ahead with the bidirectional LSTM model which gave us 93% accuracy.

B. Testing out the Pre-Trained T5 Model

Finding a good title generation model trained on pretty good dataset was done through browsing and testing out various T5 pre-trained models from the huggingface directory. After careful trial and error "czearing/article-title-generator" model [18] was chosen. This model was trained on over 100k+ medium articles and their respective titles [17]. With the T5 pre-trained model all set up and tested with sample inputst for generating titles as seen in Table II.

TABLE I

T5 PRE-TRAINED MODEL TESTING

INPUT TEXT	TITLE GENERATED
IN LIFE, THERE ARE MOMENTS THAT ETCH THEMSELVES INTO THE FABRIC OF OUR SOULS, LEAVING INDE….	THE PAIN OF SAYING GOODBYE
THE DEBATE OVER CLIMATE CHANGE POLICIES CONTINUES TO DIVI….	CLIMATE CHANGE POLICY: A CHALLENGE FOR POLITICIANS AND SCIENTISTS

C. Incorporating the Sentiment with the Title Generation

A code was written to pre-process the input from the user, in order to eliminate the special characters, unnecessary white spaces, and numbers and lowering all the characters to lowercase and so on. Then the logic for the model to predict the emotion is written, along with the logic for generating a title.

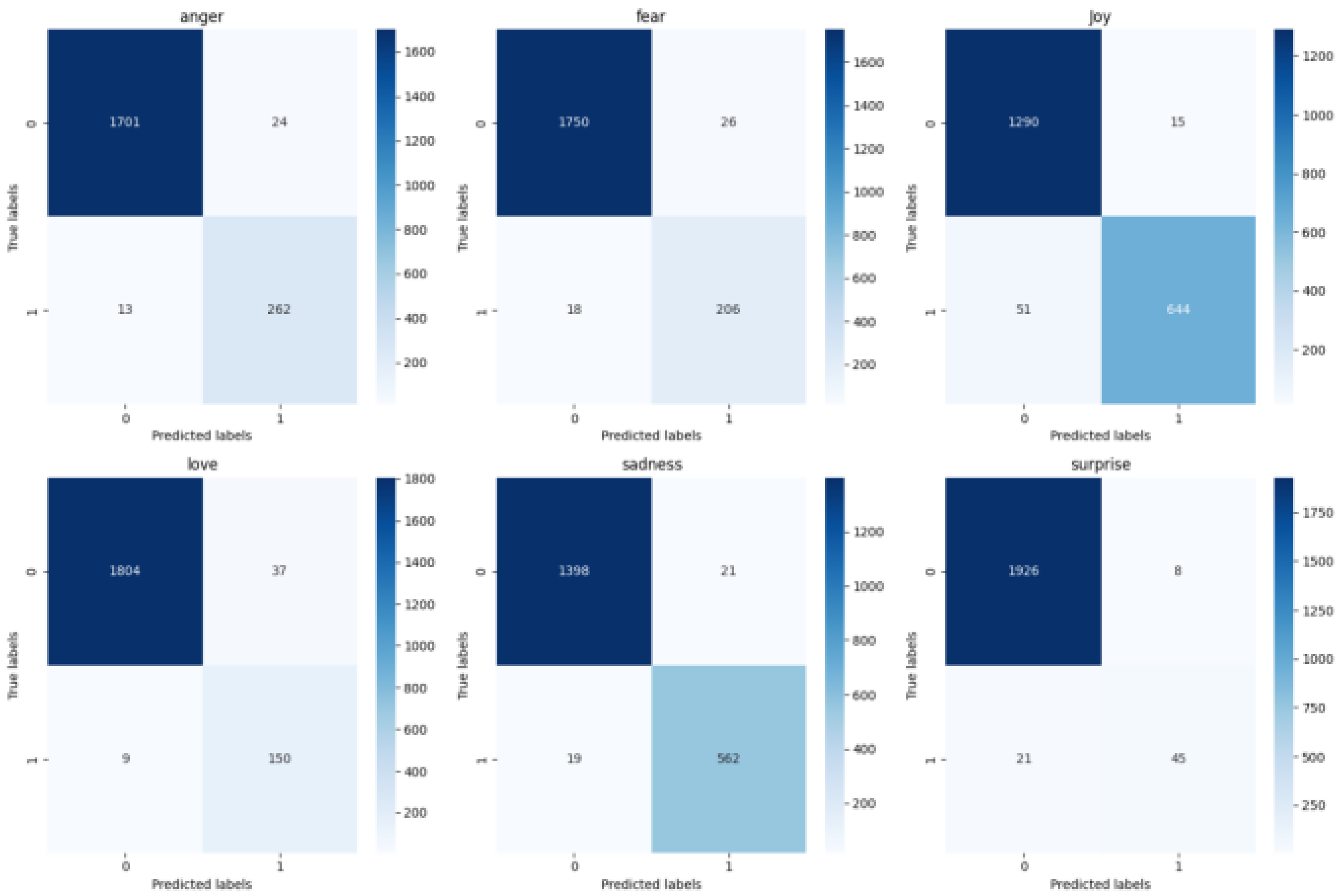

Fig. 8. Confusion Matrix of the different emotions for the sentiment analysis model.

TABLE II

SAMPLE INPUT WITH THE PREDICTED EMOTION AND THE TITLE GENERATED

Input Text	Emotion	Title Generated
The government's failure to address the rising cost of living has left many citizens feeling angry and frus….	anger	Anger and frustration in the country
In life, there are moments that etch themselves into the fabric of our souls, leav…	sadness	The Pain of Losing a Loved One

As we can see, the title is generated with the emotion being integrated in the process of generation.

V.CONCLUSION

To summarize, including sentiment analysis into title generation is a significant step forward in the field of computational linguistics. Our study demonstrates the possibility to create emotionally captivating titles that capture the essential concepts of the underlying content using sentiment analysis approach. By imbuing titles with emotional depth, we have not only increased user engagement and interest, but also provided a more meaningful browsing experience for people traversing the internet. Looking ahead, continued research and development in sentiment analysis algorithms have enormous potential to revolutionise the landscape of content discovery and interaction across many online mediums. We should expect more advancements in how users engage with and interpret digital material as we continue to explore and innovate.

REFERENCES

[1] D.R. Raman, S. Jayalakshmi, K. Arumugam, A. V. A. Raj, D. Balaji and R. Brightsingh, "Implementation of Data Analysis and Document Summarization in Social Media Data Using R and Python," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022, pp. 1457-1464, doi: 10.1109/ICIRCA54612.2022.9985479.

[2] H. Mehta, S. K. Bharti and N. Doshi, "Text summarization using Secretary problem," 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), Gunupur, Odisha, India, 2022, pp. 1-5, doi: 10.1109/iSSSC56467.2022.10051552.

[3]L. Juntong, "English Abstract Writing Based on Natural Language Generation," 2022 International Conference on Information System, Computing and Educational Technology (ICISCET), Montreal, QC, Canada, 2022, pp. 203-206, doi: 10.1109/ICISCET56785.2022.00057.

[4]M.Riku and K. Masaomi, "A title generation method with Transformer for journal articles," 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022, pp. 1115-1120, doi: 10.23919/APSIPAASC55919.2022.9979942.

[5] S. S. Kale and W. B. Andreopoulos, "Job Tailored Resume Content Generation," 2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (BigDataService), Athens, Greece, 2023, pp. 40-47, doi: 0.1109/BigDataService58306.2023.00012.

[6] T. -F. Hsieh, J. -S. Jwo, C. -H. Lee and C. -S. Lin, "Unsupervised Title Generation from Unpaired Data with N-Gram Discriminators," 2022 7th International Conference on Signal and Image Processing (ICSIP), Suzhou, China, 2022, pp. 763-766, doi: 10.1109/ICSIP55141.2022.9886556.

[7] X.Ma,J.W.Keung,X,Yu,H,Zou,J.Zang and Y,Li,"AttSum; A Deep Attention-Based Summarization Model for Bug Report Title Generation ."in IEEE Transactions on Reliability,vol.72, no.4,pp 1663-1677,Dec,2023,doi;10.1109/TR.2023.236404

[8] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[9] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pretraining,", [Online].Available:https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf, 2018.

[10]Y. Liu et al., "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[11] R. Jin and A. G. Hauptmann, "Title generation for spoken broadcast news using a training corpus," in *Proc. 6th International Conference on Spoken Language Processing (ICSLP 2000)*, 2000, pp. vol. 2, 680–683.

[12] S. Takase, J. Suzuki, N. Okazaki, T. Hirao, and M. Nagata, "Neural Headline Generation on Abstract Meaning Representation," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics,Nov. 2016, pp. 1054–1059.

[13]Gavrilov, P. Kalaidin, and V. Malykh, "Self-attentive Model for Headline Generation," in *Advances in Information Retrieval*, L. Azzopardi,B. Stein, N. Fuhr, P. Mayr, C. Hauff, and D. Hiemstra, Eds. Cham:Springer International Publishing, 2019, vol. 11438, pp. 87–93, series Title: Lecture Notes in Computer Science.

[14] P. Mishra, C. Diwan, S. Srinivasa, and G. Srinivasaraghavan, "Automatic Title Generation for Text with Pre-trained Transformer Language Model," in 2021 IEEE 15th International Conference on Semantic Computing (ICSC), Jan. 2021, pp. 17–24, iSSN: 2325-6516.

[15] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 1-135.

[16] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 267-307.

[17]Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 15-21.

[18] Tang, D., Qin, B., & Liu, T. (2015). Learning semantic representations of users and products for document level sentiment classification. *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers) (pp. 1014-1023).

[19]Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (Volume 1: Long Papers) (pp. 142-150).

[20]Praveen, 2020, "Emotions dataset for NLP", Kaggle, [Online], Available: https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp/data

[21]Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation". [Online]. Available: https://nlp.stanford.edu/pubs/glsove.pdf

[22]Meruvu Likith, 2024, "100K + Medium Articles Dataset For LLM FineTuning", [Online]. Available: https://www.kaggle.com/datasets/meruvulikith/193k-medium-articles-dataset-for-llm-finetuning.

[23]Caleb Zearing, 2022, "czearing/article-title-generator", [Online], Available: https://huggingface.co/czearing/article-title-generator.