Middle East Research Journal of Engineering and Technology
ISSN: 2789-7737 (Print) ISSN: 2958-2059 (Online)
Frequency: Bi-Monthly
DOI: https://doi.org/10.36348/merjet.2024.v04i04.001

Website: http://www.kspublisher.com/
Email: office@kspublisher.com

Harnessing Advanced NLP Techniques for Automated Personality Analysis and Future Behavior Prediction from Social Media Posts

Arman Mohammad Nakib[1*], Prottoy Khan[2], Md Mahib Ullah[3], Md Labib Kawser[4], A K M Jayed[4], Sazzad Kadir Zim[5]
[1]Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing, Jiangsu, China
[2]Computer Science and Technology, Nantong University, Nantong, Jiangsu, China
[3]Electrical Engineering and It's Automation, North China Electric Power University, Beijing, China
[4]Computer Science and Technology, Nanjing University of Information Science & Technology, Nanjing, Jiangsu, China
[5]Business Management, Nanjing University of Information Science & Technology, Nanjing, Jiangsu, China

Abstract: This work offers an integrated multitool approach that relies on state-of-the-art NLP methods for real-time text analysis, specifically in sentiment analysis, personality profiling, and knowledge graph construction. The pipeline uses abstractive summarization skills from PEGASUS model to condense long inputs from the users. That is followed by a sentiment analysis process that applies BERTs to classify the summarized text's emotional sentiment as either positive, negative, or neutral. The framework also derives personality traits from emotion and expects probable future behaviors by mapping the sentiment graph against the model defining the traits. Besides, for text preprocessing, we use the NLTK library for tokenization and removing stopwords, always extracting important keywords from users' inputs. These keywords are then used to build a knowledge graph, which is then implemented using NetworkX and Matplotlib to show connections between the identified ideas. This knowledge graph is used for generating the forecast of interconnections between the keywords to provide a clear and concise approach in comparison with the complex interconnection maps. The proposed system enables input text to be in a different language and the output summaries, sentiments, and knowledge graphs in the same required language as the input text. Combinedly, the framework intends to provide real-time, precise analysis of the contents of social media posts for future course of action prediction and for use in future applications like health, mental health checks, and analysis of social behavior.
General Terms: Chenopodium Album, Starch

Keywords: PEGASUS Model, BERT Model, NLTK toolkit, Social Media Posts Analysis, Personality and Future Prediction.

Research Paper

***Corresponding Author:**
Arman Mohammad Nakib
Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing, Jiangsu, China

How to cite this paper:
Arman Mohammad Nakib *et al* (2024). Harnessing Advanced NLP Techniques for Automated Personality Analysis and Future Behavior Prediction from Social Media Posts. *Middle East Res J. Eng. Technol*, *4*(4): 98-106.

Article History:
| Submit: 26.10.2024 |
| Accepted: 25.11.2024 |
| Published: 30.11.2024 |

1.0 INTRODUCTION

A significant quantity of data is produced with the help of online platforms in the modern era; thus, powerful algorithms should be created to process all such data. Facebook and other social networks, blog, and forums reveal to the users' thoughts and opinions, help to understand their emotional state, and even diagnose mental health issues if any. The information presented in such text cannot be easily analyzed and insights such as the sentiment behind it or the user's possible reaction should be derived using the natural language processing methods which are more advanced. This paper proposes an approach that utilizes several of the most advanced NLP models and instruments to create a tool for the analysis of the text generated by users. For the summarization of the information, the PEGASUS model has been used while BERT has been used for the sentiment analysis of the information entered, within the system NLTK has been implemented for the extraction of keywords from the input information. PEGASUS, the transformer model, designed for abstractive summarization is capable of condensing large textual input into compact summaries without much Loss of important Information. This is beneficial in avoid duplicity and makes the next sentiment analysis more efficient.

After the summary is created, the sentiment analysis is done by BERT (Bidirectional Encoder Representations from Transformers), it distinguishes the sentiment of the text as positive, negative, neutral. Thanks to bidirectional pretrained MSA BERT is highly accurate in calculating the general sentiment, considering the context from both directions, left to right, and right to left. Emotion is considered fundamental in personality traits and future behaviors and this is why sentiment analysis is considered relevant in modeling the personality traits.

Apart from sentiment analysis tokenization and music word extraction are carried out using NLTK. Tokenization divides the text into words and then filters out stopwords which can be words like “the,” “and.” These keywords are used to define basic concepts and topics most relevant to the user’s content.

Lastly, using the same libraries NetworkX and Matplotlib, we formulate a knowledge graph. Knowledge graphs described connections between significant entities: in this case, the user and the most significant extracted keywords; it is useful for improved comprehension and visual perception of the analyzed content. It also helps in the model prediction since the system predicts future behaviors or future health risks as may be associated with the extracted data.

The proposed framework is designed in such a way that the users are allowed to input data in different languages. This increases the flexibility of the system and its usefulness in actual use cases such as mental health tracking, where users might input their data in their first language. The system’s ultimate goal is to offer instantaneous records to help both normal persons and occupations to evaluate trend in mood and action. In accomplishing this integration of these advanced models, this work continues the pursuit of automated content analysis, for approaches in health diagnosis, social conduct and future behavior prognostications.

2.0 LITERATURE REVIEW

This literature review examines existing research on the application of NLP techniques to social media data, with a particular focus on sentiment analysis, topic modeling, and personality prediction. It explores the methodologies employed in these studies, the effectiveness of various models, and the implications of their findings. By understanding the current state of research, this review aims to highlight the strengths and limitations of existing approaches and identify opportunities for further advancements in the field.

2.1 Sentiment Analysis on Social Media Activities Using NLP Techniques

This is the discusses various methods for analyzing sentiments on Twitter. The authors explore machine learning techniques such as Naive Bayes, Maximum Entropy, and Support Vector Machines (SVM). They highlight the challenges in dealing with the informal language used in social media posts.

This study focuses on the application of traditional machine learning techniques for sentiment analysis on Twitter data. The researchers collected tweets and manually annotated them for sentiment. The study compares the performance of different classifiers and discusses the impact of feature selection on the accuracy of sentiment prediction (Pak, A., & Paroubek, P., 2010; Hasan, M. R., Maliha, M., & Arifuzzaman, M., 2019; Khan, R., Shrivastava, P., Kapoor, A., Tiwari, A., & Mittal, A., 2020; Kanakaraj, M., & Guddeti, R. M. R., 2015; Hossain, S. and Nur, T.I., 2024).

2.2 Topic Modeling in Social Media

This study introduces Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) as methods for topic modeling. The authors apply these techniques to Twitter data to identify prevalent topics.

The study applies LDA and NMF to model topics in a large corpus of Twitter data. By categorizing tweets into different topics, the authors demonstrate how these methods can uncover underlying themes in social media conversations. The paper provides a detailed comparison of LDA and NMF, highlighting the strengths and limitations of each approach (Blei, D. M., Ng, A. Y., & Jordan, M. I., 2003; Rohani, V. A., Shayaa, S., & Babanejaddehaki, G., 2016; Molenaar, A., Lukose, D., Brennan, L., Jenkins, E. L., & McCaffrey, T. A., 2024, Atagün, E., Hartoka, B., & Albayrak, A., 2021; HOSSAIN, S. and HENA, H., 2024).

2.3 Personality Prediction from Text

This study presents an open vocabulary approach to predict personality traits from social media text. The authors use a large dataset of Facebook status updates and apply linguistic and psychological analyses to infer personality traits.

The researchers utilize an open vocabulary technique to analyze the language used in Facebook status updates. They correlate linguistic features with personality traits, as measured by standard psychological assessments. The study shows that language use on social media can be a reliable indicator of personality traits (Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., ... & Seligman, M. E., 2014; Vora, H., Bhamare, M., & Kumar, D. K. A., 2020; Feizi-Derakhshi, A. R., Feizi-Derakhshi, M. R., Ramezani, M., Nikzad-Khasmakhi, N., Asgari-Chenaghlu, M., Akan, T., ... & Jahanbakhsh-Naghadeh, Z., 2022; Kampman, O., Barezi, E. J., Bertero, D., & Fung, P., 2018; Hossain, S., Hossen, M.S.B., Zim, S.K. and El Hebabi, I., 2024).

2.4 PEGASUS for Text Summarization

This study introduces PEGASUS, a state-of-the-art model for abstractive text summarization. The model is pre-trained on a large corpus using gap-sentences, making it highly effective for generating summaries.

The PEGASUS model is designed to handle the complexities of summarizing long documents. The authors describe the pre-training process, where the model learns to predict missing sentences in a text. The paper includes extensive evaluations, showing PEGASUS's superiority over previous models in summarization tasks (Lee, J., Dang, H., Uzuner, O., & Henry, S., 2021; Zhang, J., Zhao, Y., Saleh, M., & Liu, P., 2020; Alsuhaibani, M; Goodwin, T. R., Savery, M. E., & Demner-Fushman, D., 2020; Hossain, S., Akon, T. and Hena, H.).

2.5 BERT for Sentiment Analysis

This study introduces BERT, a transformer-based model pre-trained on a large corpus of text. The model achieves state-of-the-art results in various NLP tasks, including sentiment analysis.

BERT's bidirectional training approach enables it to understand context from both directions in a sentence. The paper details the architecture and training methodology of BERT, demonstrating its effectiveness across multiple benchmarks. The model's performance in sentiment analysis tasks highlights its capability in understanding nuanced emotional content in text (Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., 2018; Alaparthi, S., & Mishra, M., 2021; Catelli, R., Pelosi, S., & Esposito, M., 2022; Batra, H., Punn, N. S., Sonbhadra, S. K., & Agarwal, S., 2021).

2.6 Multi-Aspect Sentiment Analysis with Bert

This study explores fine-tuning BERT for multi-aspect sentiment analysis, allowing the model to assess sentiment across different aspects of a text. The study applies this approach to product reviews, demonstrating its utility in extracting nuanced sentiment information.

The researchers fine-tune BERT to handle multiple sentiment aspects within a single text, such as evaluating both the quality and price of a product in a review. The paper provides insights into the model's architecture adjustments and training process, illustrating its enhanced performance in multi-aspect sentiment tasks (Tang, T., Tang, X., & Yuan, T., 2020; Wu, Z., Ying, C., Dai, X., Huang, S., & Chen, J., 2020; Ahmed, M., Pan, S., Su, J., Cao, X., Zhang, W., Wen, B., & Liu, Y., 2022; Wankhade, M., Annavarapu, C. S. R., & Abraham, A., 2023).

These existing works collectively provide a strong foundation for understanding the application of NLP techniques in social media analysis. By leveraging advanced models like PEGASUS and BERT, this study aims to build on these findings and further enhance the accuracy and reliability of personality prediction and behavior analysis from social media posts (Tech, D. D. E. C. M., Lokesh, P. S., Mounika, N.; Datta, N., 2024; Kachhoria, R., Daga, N., Ramteke, H., Akotkar, Y., & Ghule, S., 2024; Nakib, A.M., Luo, Y., Emon, J.H. and Chowdhury, S., 2024; Nakib, A.M. and Barua, B.).

3.0 METHODOLOGY

3.1 Collecting Input from User

The first mode of operation is initiated by the user and he is expected to feed the system with his posts. The input is a set of raw user-created textual data, which can be from personal blogs or walls at for example, Facebook. These details serve as the input to other processes such as: summarization, sentiments analysis, keyword extraction, and generating a knowledge graph.

3.2 Generating Summary of the Posts by Pegasus Model

PEGASUS (Pre-training with Extracted Gap-sentences for Abstractive Summarization Sequence-to-sequence) is based on the Transformer model. The Transformer uses an encoder-decoder architecture, which can be described as follows:

The input text T={t1, t2, … , tn} is tokenized into subword tokens {x1, x2, … , xm}, where each xi represents a token from the original text sequence.

The PEGASUS encoder maps the input sequence {x1, x2, … , xm} into a sequence of hidden states H={h1, h2, … , hm}, where:
hi = Transformer Encoder (xi, Hi−1)

The encoder uses multi-head attention and feed-forward layers, where multi-head attention is computed as:
Attention (Q, K,V) = softmax (QK^T/dk)V
Here, Q (query), K (key), and V (value) are matrices derived from the input hidden states.

The PEGASUS decoder generates the summary by attending to the encoder's hidden states:
yj = Transformer Decoder (yj−1, H)
This produces the summary {y1, y2, … , yk}, where each yj represents a token in the summarized output.

3.3 SENTIMENT ANALYSIS BY BERT MODEL

BERT stands for Bidirectional Encoder Representations from Transformers) is a language representation model that works by understanding the bidirectional context of a sentence. It operates as follows:

The summarized text S={s1, s2,…, sl} is tokenized into subword tokens {x1, x2, …, xn}, and then passed to the BERT model.

BERT uses word piece embeddings for each token:
ei = E(xi) + P(i)
where E(xi) is the embedding of the token, and P(i) is the positional encoding.

1. Similar to PEGASUS, BERT uses multi-head attention and feed-forward layers. The self-attention is computed as:
   A (Q, K, V) = softmax (QK^T/dk) V

Q, K, and V are projections of the hidden states from the previous layer. This produces context-aware representations of each token.

For sentiment analysis, BERT adds a classification layer on top of the final hidden state corresponding to the [CLS] token. The [CLS] token's representation h[CLS] is used to predict the sentiment:

$p\ (\text{sentiment}|S) = \text{softmax}\ (W\cdot h[CLS] + b)$

where W and b are learned parameters.

3.4 Recognizing the Sentiment of the Condensed Text

The mood of the input that the user uses to interact with the bot is the exact same as the feeling that results from the sentiments generated by the BERT model. The system identifies if the user is in a "Very Negative" "Negative" "Neutral" "Positive" "Very Positive" Emotional State. It also has a significant function for the analysis of the personality and the prognosis for the further manifestations of the user's emotional and behavioral state.

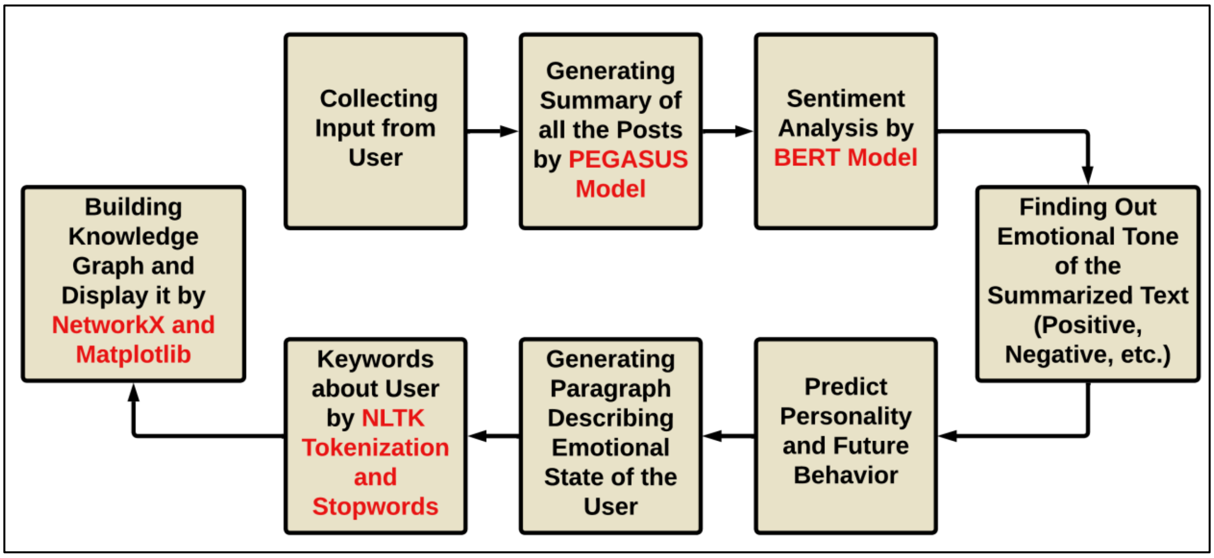

Figure 1: Proposed Design

3.5 Forecast Personality and Future Behavior

Using the determined probable emotional tone of the produced summary text, the system then assesses the user's characteristics and potentially expected future conduct. The sentiment analysis of the text by BERT is related to personality features. For instance, the positive sentiment could mean that the user is outgoing and a positive thinker while the negative sentiment would be equal to an introvert and a gloomy-user. Further, by considering this emotional tone, the system analyzes the future behavioral proclivities and suggests information such as if the user is going to have a rough time or otherwise, has a rosy future ahead.

3.6 Creating a Paragraph Expressing Pre-Determined Feelings of the User

Based on the calculated personality by the system and possibly future behavior, the system produces a descriptive paragraph of the user. This paragraph discusses the user's based on the summarized data and sentiment analysis this means and in this paragraph the user emotional analysis and personality traits are presented in form of a narrative. This allows the writer to portend the users' personality and the behaviours they conceive based on their euphony of the keys pad.

3.7 Special Keywords about User Identified by NLTK Tokenization and by Elimination of Stopwords

In order to process the user's content even more, the program analyses the summarized text for keywords. The system adopts the NLTK and splits each document into words and then eliminates the English stop words meaning less in the document. This leads to an optimization of the keywords that are both significant and significant enough for the user's posts. These keywords are then supported by the extra terms denoting user personality and opinion which provides a greater depth of analysis when generating a knowledge graph.

3.8 Create the Knowledge Graph and Represent it Using NetworkX and Matplotlib

The last one of them consists in building and visualizing a knowledge graph. In this step, the keywords identified in the previous step, with the user's personality, and sentiment will create nodes and edges for a graph. The NetworkX, that is a Python library for using and analyzing complex networks is used for making graphs, and Matplotlib is used for displaying the result. The core vertice in the graph is the user, while the other vertices are constructed from the keyword and personality traits. This makes it easier to present the user's mood, interests in specific topics and even personality in a graphical way.

4.0 ADVANCEMENTS OF THIS RESEARCH FROM EXISTING WORKS

The work in this paper distinguishes itself from existing works through several innovative aspects and unique contributions:

4.1 Integrated Pipeline for Comprehensive Analysis

Existing studies often tackle individual NLP tasks in isolation, such as sentiment analysis or topic modeling, which may overlook the interconnected nature of social media data. Conversely, the methodology of this paper offers an integrated pipeline that harmonizes multiple NLP tasks. This approach starts with post summarization using PEGASUS, capturing the essence of user content efficiently. The summarized posts then undergo personality analysis using BERT, followed by the prediction of future behaviors based on inferred personality traits. By orchestrating these tasks within a unified framework, this methodology delivers a holistic understanding of social media content, transcending the limitations of isolated analyses.

4.2 Use of Advanced Transformer Models

Traditional machine learning techniques or basic NLP models employed in prior studies may struggle to capture the complexity and subtleties inherent in social media text. In contrast, this approach harnesses the power of advanced transformer models, specifically PEGASUS for summarization and BERT for sentiment analysis and personality inference. These state-of-the-art models excel in capturing intricate linguistic patterns and contextual nuances, thereby enhancing the accuracy and depth of insights derived from social media data.

4.3 Summarization as a Preprocessing Step

Many existing studies directly analyze raw text from social media posts, potentially inundating the analysis with noise and irrelevant information. By integrating post summarization as a preprocessing step, this methodology mitigates this issue effectively. PEGASUS succinctly distills the key points of each post, ensuring that subsequent analyses focus on the most salient content. This preprocessing step not only enhances the efficiency of downstream analyses but also improves the quality of insights derived from the data by prioritizing relevant information.

4.4 Predictive Capability for Future Behavior

While personality prediction is a common focus in previous studies, few extend their analyses to predict future behaviors based on inferred personality traits. This methodology innovatively bridges this gap by linking personality traits to potential future behaviors or tendencies. This predictive aspect empowers applications in diverse domains, including marketing strategies, mental health interventions, and personalized user engagement. By leveraging the predictive power of inferred personality traits, this methodology facilitates proactive decision-making and tailored interventions, thereby maximizing the utility of social media data for practical applications.

4.5 Keyword Extraction and Visualization

Although keyword extraction and topic modeling are prevalent in some studies, their integration into a comprehensive personality and behavior analysis is rare. This methodology seamlessly incorporates keyword extraction using NLTK, enriching the analysis with key themes and topics present in the user's posts. Moreover, the visualization of these keywords within a knowledge graph offers an intuitive representation of the relationships and trends within the data. This visualization aids in uncovering underlying patterns and connections, fostering deeper insights into user behavior and preferences.

4.6 Comprehensive Data Usage

While some studies are confined to specific datasets or social media platforms, your analysis encompasses a diverse set of social media posts, ensuring the generalizability and robustness of this findings. This comprehensive data usage broadens the applicability of this study's insights across various contexts and content types, enhancing its relevance and impact in real-world scenarios. By leveraging a diverse array of social media data, this methodology delivers insights that are adaptable to different domains and use cases, thereby maximizing its practical utility and significance.

5.0 EXPERIMENT AND RESULT

The experimented results are as follows:
The code will ask:

How many posts would you like to input?
The user gave the sample input: 3

Then the user gave the three posts:
Enter post 1: I woke up today with a heart full of gratitude and realized that happiness isn’t something we have to chase – it’s something we can choose every single day. 💖 Life has its ups and downs, but right now, I’m focusing on the things that make my heart sing: good friends, beautiful moments, and the incredible opportunities all around me. This week has been full of wonderful surprises, from spontaneous get-togethers with loved ones to moments of peace and reflection. It’s in these little moments that I’m reminded how lucky I am. 🌸 Life is a gift, and it’s important to savor every part of it. To everyone out there – remember that happiness is in the now. Choose joy, appreciate the present, and spread kindness wherever you go. Here’s to a lifetime of positivity, love, and endless smiles! 💫
Enter post 2: Today has been a day filled with so many simple, beautiful moments, and I just can’t help but feel grateful. 💖 Whether it was the warm sunlight streaming through my window, the sound of my favorite song playing on the radio, or the sweet text from a friend, I realized how much joy there is in life’s everyday magic. It’s easy to get bogged down by the rush of life, but when we slow down, we start to see the beauty in the ordinary. The past few days have reminded me of how important

it is to stop and soak in these moments. 😌 I'm feeling so lighthearted and content right now, and I want to pass on that positivity to all of you! Let's take a moment to smile, breathe, and appreciate everything we have. Life's too short not to enjoy every second of it. Sending love and happiness your way! 💫 💖

Enter post 3: Life has been giving me so many reasons to smile lately, and I just had to share my joy with all of you! Sometimes, we get so caught up in chasing the next big thing that we forget to appreciate the beauty of the little things – like a warm cup of coffee in the morning, the sound of birds chirping outside, or a random act of kindness from a stranger. Today, I'm filled with gratitude for every small blessing that has come my way. Whether it's a spontaneous chat with an old friend, the laugh of a loved one, or even just the calmness of a cozy evening, I'm realizing more and more that happiness is all around us, waiting to be noticed. 🌿 💕 I'm sending positive vibes to everyone reading this. May your day be filled with laughter, love, and light! Keep shining, and remember that every moment counts. Let's celebrate the present and find joy in everything we do. 💫 💖

Then the code will give the results:

Combined Summary: Today, I want to share with you some of the things I'm grateful for in my life.

Sentiment: Positive.

Personality Prediction: Optimistic, social

Future Prediction: Positive outlook, good opportunities.

Summary Paragraph: The individual is characterized as Optimistic, social based on their expressed sentiments, which are positive. This reflects their emotional state and tendencies towards optimistic, social traits.

Extracted Keywords: ['Today', 'I', 'want', 'share', 'things', 'I', 'grateful', 'life', 'optimistic', 'positive']

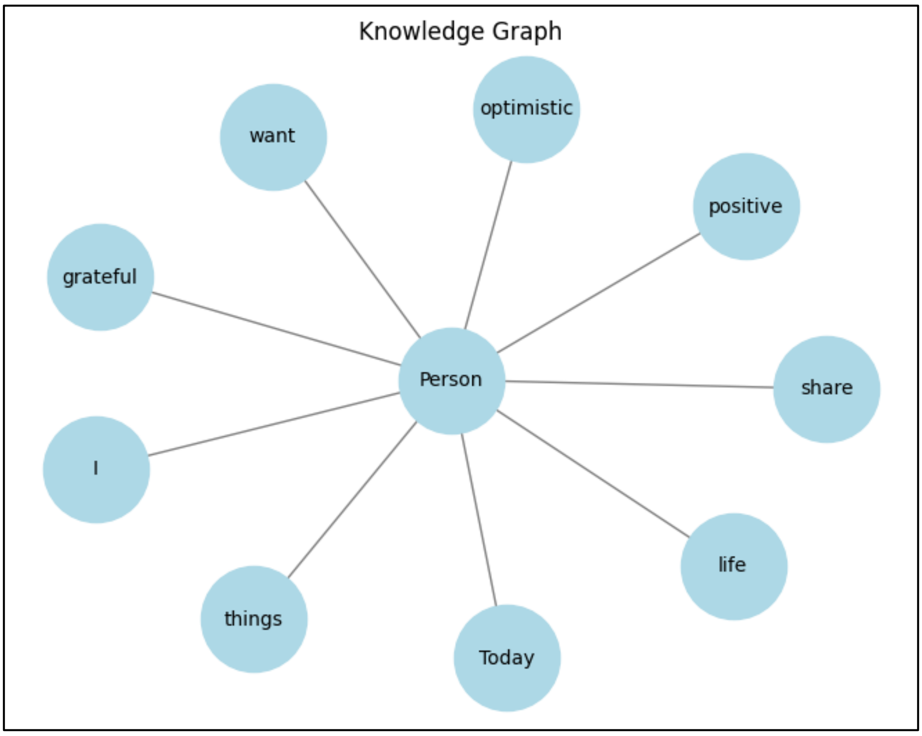

Figure 2: Knowledge Graph

6.0 ADVANTAGES

By leveraging advanced natural language processing (NLP) techniques and pre-trained models, the methodology facilitates automated analysis of social media content. This automation drastically reduces the time and effort required for manual analysis, enabling researchers to process large volumes of data efficiently. Automated analysis ensures consistent results across datasets, minimizing the potential for human errors and subjective biases inherent in manual interpretation. Moreover, it allows for real-time processing of data streams, enabling timely insights into user behavior and sentiment dynamics on social media platforms.

The methodology adopts a multi-faceted approach by integrating text summarization, sentiment analysis, and keyword extraction, thereby providing a comprehensive understanding of the user's personality and behavior. Summarization condenses lengthy posts into concise summaries, capturing the essence of the content. Sentiment analysis offers valuable insights into the user's emotional state, discerning sentiments such as happiness, sadness, or excitement expressed in the posts. Additionally, keyword extraction identifies recurring themes and topics, shedding light on the user's interests and concerns. By combining these techniques, the methodology generates rich insights into various aspects of the user's personality, preferences, and sentiments.

7.0 LIMITATIONS

It may be argued that training data for pre-existing models contains prejudice, therefore, personality inferences may be skewed as well. Eradicating these biases call for assessment and minimization of such bias then the improvement of the given model further will improve fairness.

The process of analyzing data for social media also provides a privacy issue in a sense that anonymization of data may not be sufficient enough to protect users' information. The principles that should be followed are confidentiality of data handling, obtaining consent from subjects, encryption, and ethical conduct.

8.0 CONCLUSION

In conclusion, the methodology presented in this study represents a significant advancement in the field of social media analysis, particularly in understanding user behavior and personality traits through Facebook posts. By leveraging state-of-the-art natural language processing (NLP) models and techniques, including text summarization, sentiment analysis, keyword extraction, and knowledge graph visualization, the study offers a robust framework for comprehensive analysis.

The integration of multiple NLP methods allows for a nuanced understanding of the user's personality, preferences, and emotional states. Text summarization condenses the rich content of Facebook posts into succinct summaries, capturing the essence of the user's thoughts and experiences. Sentiment analysis provides insights into the user's emotional tendencies, helping identify patterns of positivity, negativity, or neutrality in their expressions. Keyword extraction further enriches the analysis by highlighting recurring themes and topics of interest, offering valuable context for understanding the user's motivations and concerns.

Moreover, the visualization of findings through a knowledge graph enhances the interpretability and accessibility of the results. The knowledge graph visually represents the relationships between the user's personality traits and the key descriptors extracted from their posts, providing a clear and intuitive overview of their personality profile. This visual representation facilitates easier interpretation of the data, enabling researchers and stakeholders to grasp the user's characteristics and tendencies at a glance.

Overall, the study demonstrates the potential of advanced NLP techniques in unlocking valuable insights from social media data. By combining automated analysis with comprehensive insights and intuitive visualization, the methodology presented offers a powerful tool for understanding user behavior and informing various applications, including personalized marketing, mental health monitoring, and social research. As NLP continues to advance, future research may further refine and expand upon this methodology, unlocking even deeper insights into human behavior in the digital age.

9.0 FUTURE WORK

In the future work section, several avenues for further research and development are identified, aiming to build upon the foundations laid by the current study and enhance the methodology's effectiveness and applicability.

Fine-tuning the PEGASUS and BERT models using additional data specific to the context of social media posts could improve their accuracy and relevance in summarization and sentiment analysis tasks. This may involve collecting and annotating a large dataset of Facebook posts to fine-tune the models to better capture the nuances of user-generated content.

Conducting thorough bias analyses and mitigation strategies to reduce biases present in the pre-trained models is essential. This involves identifying and addressing biases related to gender, race, culture, and other demographic factors that may influence the models' predictions and inferences.

Adapting the methodology to analyze Facebook posts in languages other than English can broaden its applicability and enable cross-cultural insights. This involves training and fine-tuning NLP models on multilingual datasets to ensure accurate and reliable analysis across different linguistic contexts.

Developing systems and algorithms to analyze Facebook posts in real-time allows for timely insights and interventions. This may involve building scalable and efficient pipelines for data ingestion, processing, analysis, and visualization, capable of handling large volumes of social media data in real-time.

Establishing guidelines and best practices for the responsible use of NLP techniques in analyzing social media data is crucial. This involves addressing privacy concerns, ensuring informed consent, protecting user data, and mitigating potential harms such as misinformation, algorithmic biases, and invasions of privacy.

REFERENCES

- Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).
- Hasan, M. R., Maliha, M., & Arifuzzaman, M. (2019, July). Sentiment analysis with NLP on Twitter data. In *2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2)* (pp. 1-4). IEEE.
- Khan, R., Shrivastava, P., Kapoor, A., Tiwari, A., & Mittal, A. (2020). Social media analysis with AI:

sentiment analysis techniques for the analysis of twitter covid-19 data. *J. Crit. Rev*, *7*(9), 2761-2774.

- Kanakaraj, M., & Guddeti, R. M. R. (2015, February). Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. In *Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015)* (pp. 169-170). IEEE.
- Hossain, S. and Nur, T.I., 2024. Gear up for safety: Investing in a new automotive future in China. *Finance & Accounting Research Journal*, *6*(5), pp.731-746.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.
- Rohani, V. A., Shayaa, S., & Babanejaddehaki, G. (2016, August). Topic modeling for social media content: A practical approach. In *2016 3rd international conference on computer and information sciences (ICCOINS)* (pp. 397-402). IEEE.
- Molenaar, A., Lukose, D., Brennan, L., Jenkins, E. L., & McCaffrey, T. A. (2024). Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study. *Journal of Medical Internet Research*, *26*, e47826.
- Atagün, E., Hartoka, B., & Albayrak, A. (2021, September). Topic modeling using LDA and BERT techniques: Teknofest example. In *2021 6th International Conference on Computer Science and Engineering (UBMK)* (pp. 660-664). IEEE.
- HOSSAIN, S., & HENA, H. (2024). The study explores the correlation between cultural influence and community engagement in fostering social safety for tourists.
- Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., ... & Seligman, M. E. (2014). The online social self: An open vocabulary approach to personality. *Assessment*, *21*(2), 158-169.
- Vora, H., Bhamare, M., & Kumar, D. K. A. (2020). Personality prediction from social media text: An overview. *Int. J. Eng. Res*, *9*(05), 352-357.
- Feizi-Derakhshi, A. R., Feizi-Derakhshi, M. R., Ramezani, M., Nikzad-Khasmakhi, N., Asgari-Chenaghlu, M., Akan, T., ... & Jahanbakhsh-Naghadeh, Z. (2022). Text-based automatic personality prediction: A bibliographic review. *Journal of Computational Social Science*, *5*(2), 1555-1593.
- Kampman, O., Barezi, E. J., Bertero, D., & Fung, P. (2018). Investigating audio, visual, and text fusion methods for end-to-end automatic personality prediction. *arXiv preprint arXiv:1805.00705*.
- Hossain, S., Hossen, M. S. B., Zim, S. K., & El Hebabi, I. (2024). Cultural dynamics and consumer behavior: An in-depth analysis of Chinese preferences for western imported products. *GSC Advanced Research and Reviews, 20*(3), 158-167.
- Lee, J., Dang, H., Uzuner, O., & Henry, S. (2021, June). MNLP at MEDIQA 2021: fine-tuning Pegasus for consumer health question summarization. In *Proceedings of the 20th Workshop on Biomedical Language Processing* (pp. 320-327).
- Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020, November). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International conference on machine learning* (pp. 11328-11339). PMLR.
- Alsuhaibani, M. Fine-Tuned PEGASUS: Exploring the Performance of the Transformer-Based Model on a Diverse Text Summarization Dataset.
- Goodwin, T. R., Savery, M. E., & Demner-Fushman, D. (2020, December). Flight of the PEGASUS? Comparing transformers on few-shot and zero-shot multi-document abstractive summarization. In *Proceedings of COLING. International Conference on Computational Linguistics* (Vol. 2020, p. 5640). NIH Public Access.
- Hossain, S., Akon, T. and Hena, H., Do creative companies pay higher wages? Micro-level evidence from Bangladesh.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. *Journal of Marketing Analytics*, *9*(2), 118-126.
- Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics*, *11*(3), 374.
- Batra, H., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2021). Bert-based sentiment analysis: A software engineering perspective. In *Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I 32* (pp. 138-148). Springer International Publishing.
- Tang, T., Tang, X., & Yuan, T. (2020). Fine-tuning BERT for multi-label sentiment analysis in unbalanced code-switching text. *IEEE Access*, *8*, 193248-193256.
- Wu, Z., Ying, C., Dai, X., Huang, S., & Chen, J. (2020). Transformer-based multi-aspect modeling for multi-aspect multi-sentiment analysis. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II 9* (pp. 546-557). Springer International Publishing.
- Ahmed, M., Pan, S., Su, J., Cao, X., Zhang, W., Wen, B., & Liu, Y. (2022). BERT-ASC: Implicit

Aspect Representation Learning through Auxiliary-Sentence Construction for Sentiment Analysis. *arXiv preprint arXiv:2203.11702*.

- Wankhade, M., Annavarapu, C. S. R., & Abraham, A. (2023). MAPA BiLSTM-BERT: multi-aspects position aware attention for aspect level sentiment analysis. *The Journal of Supercomputing*, *79*(10), 11452-11477.
- Tech, D. D. E. C. M., Lokesh, P. S., Mounika, N., Gopichand, V., & Kumar, P. B. ADVANCMENTS IN TEXT SUMMARIZATION AND EXTRACTIVE QUESTION-ANSWERING: A MACHINE LEARNING APPROACH.
- Datta, N. (2024, February). Extractive Text Summarization of Clinical Text Using Deep Learning Models. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-6). IEEE.
- Kachhoria, R., Daga, N., Ramteke, H., Akotkar, Y., & Ghule, S. (2024, March). Minutes of Meeting Generation for Online Meetings Using NLP & ML Techniques. In *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 1-6). IEEE.
- Nakib, A.M., Luo, Y., Emon, J.H. and Chowdhury, S., 2024. Machine learning-based water requirement forecast and automated water distribution control system. *Computer Science & IT Research Journal*, *5*(6), pp.1453-1468.
- Nakib, A.M. and Barua, B., AQUA FLOW MASTER: INTELLIGENT LIQUID FLOW CONTROL AND MONITORING SYSTEM.