Sentiment Analysis of ChatGPT Tweets Using Transformer Algorithms

Sugeng Winardi[1], **Mohammad Diqi**[2*], **Arum Kurnia Sulistyowati**[3], **Jelina Imlabla**[4]
[1,3,4] Dept. of Information System, Universitas Respati Yogyakarta
[2] Dept. of Informatics, Universitas Respati Yogyakarta
[*]Email: diqi@respati.ac.id

Abstract

This study explores the application of the Transformer model in sentiment analysis of tweets generated by ChatGPT. We used a Kaggle dataset consisting of 217,623 instances labeled as "Good", "Bad", and "Neutral". The Transformer model demonstrated high accuracy (90%) in classifying sentiments, particularly in predicting "Bad" tweets. However, it showed slightly lower performance for the "Good" and "Neutral" categories, indicating areas for future research and model refinement. Our findings contribute to the growing body of evidence supporting deep learning methods in sentiment analysis and underscore the potential of AI models like Transformers in handling complex natural language processing tasks. This study broadens the scope for AI applications in social media sentiment analysis.

Keywords*: Sentiment Analysis, Transformer Model, ChatGPT, Deep Learning, Social Media Analysis*

INTRODUCTION

Sentiment analysis, also known as opinion mining, is a prominent field of study in natural language processing that focuses on identifying, extracting, and quantifying subjective information from source materials (Hertina et al., 2021). Its significant role can be primarily attributed to its ability to parse large volumes of unstructured data, such as social media posts, online reviews, or blog entries, and distill this into actionable insights relating to public opinion or emotional sentiment (Mudinas et al., 2018). Particularly in social media platforms like Twitter, sentiment analysis has proven instrumental in various applications, such as brand monitoring, market trend prediction, and even understanding socio-political climates (Pilař et al., 2021). These platforms generate a vast amount of data every second, and sentiment analysis provides a mechanism to sift through this 'big data' and gauge public sentiment on various topics, making it an indispensable tool in the modern digital era.

As an integral part of artificial intelligence, machine learning has significantly contributed to sentiment analysis in recent years. Models such as Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbors (KNN), and XGBoost have been used extensively due to their strong predictive performance (Preto et al., 2022). For instance, with its high dimensional feature handling capability, SVM has been proven effective in sentiment classification tasks, particularly in the case of linearly separable data (Kulkarni et al., 2021). On the other hand, Random Forests, with their ensemble approach of multiple decision trees, offer high accuracy and control over-fitting (Ahmed et al., 2022). KNN, despite its simplicity, has shown commendable performance in sentiment classification by leveraging the principle of feature similarity (Shafique & Marchán, 2022). The gradient-boosting algorithm XGBoost is a recent addition to this toolbox, but with its superior speed and performance, it has rapidly gained popularity in tackling sentiment analysis tasks (Wei et al., 2020). Thus, these machine learning models have played a pivotal role in augmenting the efficacy and potential of sentiment analysis.

The advent of deep learning methodologies has had a transformative effect on sentiment analysis, providing nuanced understandings of textual data that surpass traditional machine learning techniques (Kim et al., 2018). Convolutional Neural Networks (CNNs), primarily known for their success in image processing, have also shown promising results in natural language processing tasks, capturing local dependencies within the text through their convolutional layers (Kim et al., 2018). Recurrent Neural Networks (RNNs) have emerged as a highly effective tool for sentiment analysis owing to their inherent ability to handle sequential data, making them apt for dealing with the temporal dynamics of textual data (Picozzi & Iaccarino, 2021). Particularly, Long Short-Term Memory (LSTM) units, a special kind of RNN, have demonstrated exceptional performance in sentiment analysis due to their ability to remember long-term dependencies and overcome the issue of vanishing gradients, which is common in traditional RNNs (W. Li & B. Xu, 2020). Therefore, deep learning techniques like CNNs, RNNs, and LSTMs have broadened sentiment analysis capabilities, enabling more sophisticated and accurate interpretation of sentiments in text data.

The Transformer model has revolutionized the field of sentiment analysis with its distinct architecture that deviates from conventional recurrent neural network designs (Borgman et al., 2022). The Transformer model harnesses the power of self-attention mechanisms, enabling it to weigh the relevance of words in a text sequence irrespective of their positions. This positional independence enables it to efficiently capture long-range dependencies in text, which has been a longstanding challenge in sentiment analysis (Wang & Wu, 2018). Transformer's parallel computation capacity also allows faster training times, making it highly scalable for large datasets (Ganesh et al., 2021). Its successful derivative models, like BERT and GPT, have set new benchmarks in sentiment analysis tasks, further underscoring the effectiveness of the Transformer architecture in capturing and understanding the subtle nuances of sentiment in textual data (J. Zheng & L. Zheng, 2019). Hence, the Transformer model has emerged as a powerful tool in sentiment analysis, enhancing sentiment detection's depth, scalability, and efficiency.

The primary aim of this research is to rigorously evaluate the effectiveness of Transformer algorithms in performing sentiment analysis on ChatGPT-generated tweets (Hertina et al., 2021). This exploration is grounded in the premise that understanding the sentiment implications in the outputs of language models such as GPT is essential for further enhancing and deploying these models in sensitive areas such as social media (Alomari et al., 2021). The main contribution of this study is to extend the application of Transformer models beyond conventional human-generated text to the synthetic text generated by a powerful language model, thereby expanding the applicability domain of these models.

METHODS

2.1 Dataset Description

The dataset for this research was sourced from Kaggle, a widely-recognized platform known for hosting diverse datasets for machine learning and data science applications. The dataset has 217,623 instances and is a significant collection of ChatGPT-generated tweets, which provides a rich resource for evaluating sentiment analysis techniques. The data attributes include the tweet content and corresponding labels indicating sentiment. The labels fall into three categories: 'bad', accounting for 106,695 instances; 'good', representing 55,754 instances; and 'neutral', comprising 55,174 instances. ChatGPT, a state-of-the-art language model developed by OpenAI, generated these tweets, reflecting a wide range of conversational scenarios. The tweets were labeled based on sentiment analysis algorithms, assuming that ChatGPT mimics human-like conversations accurately enough to reflect a genuine sentiment in each generated tweet. Thus, the dataset presents ample data for the study and a balanced representation across different sentiment classes, enhancing the generalizability of the results.

2.2 Data Preprocessing

Before implementing the Transformer algorithms on the dataset, it was essential to carry out a series of data preprocessing and cleaning steps to ensure the data was in an optimal state for analysis. First, duplicate entries within the dataset were identified and removed to prevent bias in the model training phase. This process was important to uphold the integrity of our experimental design, as repeated instances could inadvertently influence the model's learning. Moreover, instances with missing values were examined. Given that our dataset was comprised solely of textual data and labels, missing values were minimal. However, instances with missing labels or tweet content were removed to maintain a consistent data structure. As the dataset is textual and categorical, the concept of outliers does not apply as it does for numerical data. However, we reviewed the data to identify any instances that may fall outside the anticipated range, such as tweets that were excessively long or contained non-standard characters. These entries were suitably handled to ensure they did not skew the model's learning process. Through these preprocessing steps, we aimed to create a robust, clean dataset that would accurately reflect the performance of the Transformer algorithms in sentiment analysis.

2.3. Data Split

To ensure the robustness of our model and facilitate a fair evaluation of its performance, we employed a stratified split of the dataset into training and testing sets. Stratification ensures that the distribution of sentiment classes (good, bad, and neutral) in our splits mirrors the distribution in the original dataset, thus avoiding any potential bias. The dataset was split into 80% for training and 20% for testing. This split ratio is a commonly accepted practice in machine learning and is often chosen because it provides sufficient data for training the model while leaving an adequate portion for testing. The training set, therefore, comprised 85,356 bad, 44,604 good, and 44,140 neutral instances. The testing set consisted of 21,339 bad, 11,150 good, and 11,034 neutral instances. This division ensures that the model is trained on a substantial portion of the data, increasing its ability to generalize while preserving a significant number of instances for an unbiased evaluation of its performance on unseen data.

2.4 Transformer Model

In the sentiment analysis task of ChatGPT tweets, the Transformer algorithm is applied to the sequence of word embeddings from each tweet to predict the sentiment label. Here's a detailed explanation of how the mathematics of the Transformer algorithm works in this context.

Tokenization & Embedding: Each tweet is first tokenized into words, and each word is converted into a vector representation or embedding. For a tweet with N words, the input to the Transformer is thus an $N \times d_{model}$ matrix, where d_{model} is the dimension of the word embeddings.

Self-attention: In the self-attention mechanism, each input word is assigned a score that indicates how much it should contribute to encoding each other word. These scores are computed as follows:

- Each word is projected to a query q, a key k, and a value v vector using learned linear transformations. The dimensions of these vectors are d_{query}, d_{key}, and d_{value}, respectively.
- For each pair of words i and j, an attention score a_{ij} is computed as the dot product of their query and key vectors, followed by a softmax operation:

$$a_{ij} = \frac{exp\left(q_i^T k_j\right)}{\sum_{n=1}^{N} exp\left(q_i^T k_n\right)} \quad (1)$$

- The output vector of each word is a weighted sum of all value vectors, with the attention scores serving as weights:

$$y_i = \sum_{j=1}^{N} a_{ij}\, v_j \quad (2)$$

Multi-head attention: The self-attention mechanism is applied multiple times (H times) in parallel to capture different types of information in the data. Each parallel application is known as a "head". The output vectors of all heads are concatenated and then linearly transformed to yield the final output of the multi-head attention mechanism:

$$\begin{aligned} &MultiHead(Q, K, V) \\ &= Concat(head_1, \dots, head_H)W_O \end{aligned} \quad (3)$$

where each $head_i$ is computed as:

$$head_i = Attention\left(QW_{Qi}, KW_{Ki}, VW_{Vi}\right) \quad (4)$$

Position-wise feed-forward networks: Each word's output from the multi-head attention goes through a feed-forward neural network. This network is applied identically to each position:

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (5)$$

Sentiment Prediction: The final layer of the Transformer model is a linear layer followed by a softmax function that converts the encoding of the tweet into a probability distribution over the sentiment labels (good, bad, neutral). The predicted sentiment is the one with the highest probability.

Loss Calculation: The model is trained to minimize the cross-entropy loss between the predicted and true sentiments:

$$L = -\frac{1}{N}\sum_{n=1}^{N} y_n log(\widehat{y_n}) \quad (6)$$

Here, y_n is the true sentiment of the n^{th} tweet, and $\widehat{y_n}$ is the predicted sentiment.

2.5. Evaluation Metrics

Here are some related equations used when calculating metrics in classification tasks such as sentiment analysis:

Accuracy - The ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision (also called Positive Predictive Value) - The ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Recall (Sensitivity, Hit Rate, or True Positive Rate) - The ratio of correctly predicted positive observations to all observations in the actual class.

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

F1 Score - The weighted average of Precision and Recall. This score tries to find the balance between precision and recall.

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (10)$$

where TP: True Positive, TN: True Negative, FP: False Positive, and FN: False Negative.

RESULTS AND DISCUSSION

3.1 Model Performance

The confusion matrix, as shown in Figure 1, provides a detailed picture of the Transformer model's performance in classifying the sentiments of the ChatGPT tweets. It comprises three sentiment classes: "Bad", "Good", and "Neutral", represented in the rows and columns.

The diagonal entries of the matrix, 19631 for "Bad", 9812 for "Good", and 9930 for "Neutral", denote the true positive rates, indicating the number of instances where the model correctly identified these sentiments. These figures are quite impressive, demonstrating that the model can accurately classify tweets' sentiments most of the time.

However, there are also non-zero off-diagonal entries, representing instances where the model's predictions were inaccurate. For example, the model incorrectly predicted 780 "Good" and 442 "Neutral" tweets as "Bad", while 1066 "Bad" and 662 "Neutral" tweets were misclassified as "Good". Similarly, 642 "Bad" and 558 "Good" tweets were wrongly predicted as "Neutral".

Although these misclassifications are significantly fewer than the correct predictions, they reveal certain challenges in distinguishing between different sentiment classes. Particularly, they highlight the model's slight difficulty in accurately predicting the "Good" and "Neutral" sentiments compared to the "Bad" sentiment. Understanding and addressing these misclassifications could be a fruitful direction for future model refinement and research.

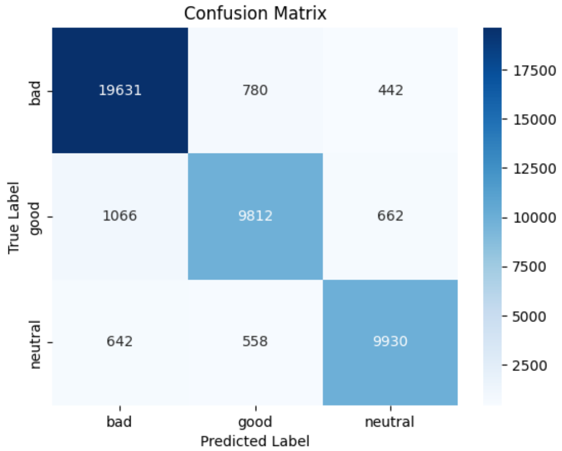

Figure 1. Confusion Matrix

The classification report, as shown in Table 1, provides a more detailed view of the Transformer model's performance across the different sentiment classes: "Bad", "Good", and "Neutral". Looking at the precision scores, the model has a high degree of exactness across all sentiment classes with 0.92 for "Bad", 0.88 for "Good", and 0.90 for "Neutral". These scores indicate that it is highly likely to be correct when the model predicts a specific sentiment.

The recall scores, which are 0.94 for "Bad", 0.85 for "Good", and 0.89 for "Neutral", provide insights into the model's completeness. The "Bad" class has the highest recall, meaning the model correctly identified 94% of the "Bad" tweets. However, the model was slightly less successful in identifying "Good" and "Neutral" tweets, with recalls of 0.85 and 0.89, respectively.

The F1-score, which is the harmonic mean of precision and recall, is a measure of the model's balanced performance. The model has fairly high F1 scores of 0.93, 0.86, and 0.90 for "Bad", "Good", and "Neutral", respectively, indicating a good balance between precision and recall.

The model's overall accuracy is 0.90, which suggests that the model correctly predicts the sentiment of a tweet 90% of the time. Both the macro-average and weighted-average scores across precision, recall, and F1-score also align around 0.90, indicating a consistent performance across the different sentiment classes.

Table 1. Classification Report

	precision	recall	f1-score
bad	0.92	0.94	0.93
good	0.88	0.85	0.86
neutral	0.90	0.89	0.90
accuracy			0.90
macro avg	0.90	0.89	0.90
weighted avg	0.90	0.90	0.90

These results demonstrate the model's effectiveness in sentiment classification. However, they highlight areas where the model could be improved, particularly in predicting "Good" and "Neutral" tweets. This underscores the need for further fine-tuning and research to optimize the model's performance across all sentiment classes.

3.2 Class-specific Analysis

In terms of individual class performance, the Transformer model showcased varying effectiveness. The "Bad" sentiment class had the highest precision, recall, and F1-score, indicating that the model was most proficient in accurately predicting this class. With a precision of 0.92, this implies that when the model predicted a tweet to have a "Bad" sentiment, it was correct 92% of the time. Furthermore, a recall of 0.94 suggests that out of all actual "Bad" tweets, the model identified 94% of them correctly. These statistics collectively contribute to a high F1-score of 0.93, suggesting a well-balanced predictive power for the "Bad" sentiment class.

The "Good" and "Neutral" classes presented more challenges for the model. Though the precision and recall metrics for these classes were still relatively high (0.88 and 0.90 for "Good"; 0.90 and 0.89 for "Neutral"), their F1-scores were lower than that of the "Bad" class, 0.86 and 0.90, respectively. This indicates that the model had a slightly harder time correctly classifying these sentiment classes. This could be due to a variety of factors. For instance, the language used in "Good" and "Neutral" tweets might be more nuanced or ambiguous, or there could be class imbalance issues in the training dataset that made the model more attuned to the "Bad" sentiment. These findings underscore the importance of further investigating and addressing these challenges to improve the model's performance across all sentiment classes.

The confusion matrix visually represents these performance metrics, allowing an intuitive understanding of how often the model made correct and incorrect predictions for each sentiment class. For example, the model correctly classified 19,631 "Bad" tweets but misclassified 780 as "Good" and 442 as "Neutral". Similarly, 9812 "Good" tweets were correctly identified, with 1066 and 662 misclassified as "Bad" and "Neutral", respectively. Lastly, the model correctly labeled 9930 "Neutral" tweets, while 642 and 558 were incorrectly labeled as "Bad" and "Good". This visual tool allows us to easily spot where the model is excelling and where it is making mistakes, thereby informing potential strategies for improvement.

3.3 Interpretation and Implications

The results of this study hold significant implications for the field of sentiment analysis, particularly regarding social media data. Achieving an overall accuracy of 90% with the Transformer model signifies a high level of performance in sentiment classification, an accomplishment that contributes to the growing body of evidence supporting deep learning-based methods for this task. Importantly, the detailed evaluation of the model's performance across the three sentiment classes - "Good", "Bad", and "Neutral" - helps identify specific areas where the model excels and where further improvements can be made. For instance, the relatively lower F1-scores for the "Good" and "Neutral" classes suggest that further research and model fine-tuning could benefit these areas. This nuanced performance analysis is crucial for developing more robust and versatile sentiment analysis models that handle various data and contexts.

From the perspective of AI applications in social media analysis, this study's findings underscore the potential of Transformer-based models for effective sentiment classification. Considering the vast amount of user-generated content on social media platforms, having an accurate and efficient tool like the Transformer model can be invaluable. For instance, businesses can use this tool to monitor public sentiment about their products or services, policymakers can gauge public opinion on various issues, and researchers can track societal trends and patterns. Moreover, the study demonstrates that AI models like ChatGPT can be valuable data sources for sentiment analysis, further expanding the scope of potential applications in social media analysis. However, it also emphasizes the need to consider potential challenges and biases in AI-generated data, underlining the importance of transparency and ethical considerations in using such models.

3.4 Challenges and Unexpected Outcomes

Throughout this study, several challenges and unexpected outcomes were encountered. A significant hurdle was dealing with human language's inherent complexity and nuance in the tweets. Sentiment analysis is not a straightforward task as it requires understanding context, detecting sarcasm, and identifying subtleties in language, which can be challenging for AI models. This complexity was particularly noticeable in the model's slightly lower performance in predicting "Good" and "Neutral" sentiments, potentially due to the greater ambiguity and subtlety in these classes compared to the more explicit language typically associated with "Bad" sentiment.

Another challenge encountered was related to the quality of the AI-generated data. Despite ChatGPT being a powerful language model, there were instances where the generated tweets were somewhat nonsensical or lacked a clear sentiment, which might have contributed to noise in the data and affected the model's performance. Furthermore, despite our rigorous data preprocessing steps, there may have been some unanticipated biases or discrepancies in the data, which could have influenced the results.

Lastly, ensuring the robustness and generalizability of the Transformer model was a challenge. While the model showed strong performance in our specific test set, there is always a concern about how it might perform with different data distributions or when confronted with novel, unseen data. Despite these challenges, the outcomes of this research provide valuable insights and guidance for further improvements in applying Transformer algorithms for sentiment analysis tasks.

3.5 Suggestions for Future Research

Given the findings of this study, future research directions could encompass several aspects. To start, further research could focus on refining and improving the Transformer model's performance for sentiment classification, particularly for the "Good" and "Neutral" classes. This could involve using larger or more diverse training datasets, incorporating additional linguistic features into the model, or applying techniques like class balancing or data augmentation to address potential biases or imbalances in the training data. In addition, while the Transformer model demonstrated a high level of performance in our study, comparing its performance with other advanced deep learning models like BERT or GPT-3 could yield informative insights.

Moreover, the application of the Transformer model could be expanded beyond the sentiment analysis of tweets. Considering its potential, it could be utilized in other areas of natural language processing, such as text summarization, machine translation, or question-answering systems. Finally, future research could also delve into the ethical and societal implications of using AI models like ChatGPT for generating social media content, providing important perspectives on the responsible and transparent use of such technologies in sentiment analysis and other AI applications.

CONCLUSION

In conclusion, this study has shed light on the potential of the Transformer model for sentiment analysis in ChatGPT-generated tweets. The model demonstrated a commendable overall accuracy of 90%, exemplifying its proficiency in deciphering sentiments within social media texts. While the Transformer model excelled in accurately categorizing "Bad" sentiments, its performance with "Good" and "Neutral" sentiments indicated certain intricacies that warrant further investigation. This research contributes to the burgeoning field of sentiment analysis by showcasing the capabilities of advanced deep learning architectures in handling nuanced sentiment classifications.

As we move forward, future research could focus on enhancing the model's performance in discerning "Good" and

"Neutral" sentiments through data augmentation techniques or by incorporating additional linguistic features. Moreover, exploring hybrid models that combine the strengths of Transformer architectures with other cutting-edge approaches might offer avenues for improved sentiment classification accuracy. Furthermore, the ethical implications of employing AI-generated data, as seen in ChatGPT tweets, call for extensive examination. Addressing these challenges will be pivotal in responsibly harnessing the capabilities of AI models for sentiment analysis within the dynamic landscape of social media discourse. Through these prospective endeavors, the full potential of Transformer algorithms in sentiment analysis could be harnessed, yielding more accurate and insightful outcomes for real-world applications.

REFERENCES

Ahmed, S. A., MonaLisa, Hussain, M., & Khan, Z. U. (2022). Supervised machine learning for predicting shear sonic log (DTS) and volumes of petrophysical and elastic attributes, Kadanwari Gas Field, Pakistan. *Frontiers in Earth Science*, *10*. https://www.frontiersin.org/articles/10.3389/feart.2022.919130

Alomari, E., Katib, I., Albeshri, A., & Mehmood, R. (2021). COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning. *International Journal of Environmental Research and Public Health*, *18*(1). https://doi.org/10.3390/ijerph18010282

Borgman, J., Stark, K., Carson, J., & Hauser, L. (2022). Deep Learning Encoding for Rapid Sequence Identification on Microbiome Data. *Frontiers in Bioinformatics*, *2*. https://www.frontiersin.org/articles/10.3389/fbinf.2022.871256

Ganesh, P., Chen, Y., Lou, X., Khan, M. A., Yang, Y., Sajjad, H., Nakov, P., Chen, D., & Winslett, M. (2021). Compressing Large-Scale Transformer-Based Models: A Case Study on BERT. *Transactions of the Association for Computational Linguistics*, *9*, 1061–1080. https://doi.org/10.1162/tacl_a_00413

Hertina, H., Nurwahid, M., Haswir, H., Sayuti, H., Darwis, A., Rahman, M., Yendra, R., & Hamzah, M. L. (2021). Data mining applied about polygamy using sentiment analysis on Twitters in Indonesian perception. *Bulletin of Electrical Engineering and Informatics*, *10*(4), 2231–2236. https://doi.org/10.11591/eei.v10i4.2325

J. Zheng & L. Zheng. (2019). A Hybrid Bidirectional Recurrent Convolutional Neural Network Attention-Based Model for Text Classification. *IEEE Access*, *7*, 106673–106685. https://doi.org/10.1109/ACCESS.2019.2932619

Kim, A. Y., Ha, J. G., Choi, H., & Moon, H. (2018). Automated Text Analysis Based on Skip-Gram Model for Food Evaluation in Predicting Consumer Acceptance. *Computational Intelligence and Neuroscience*, *2018*, 9293437. https://doi.org/10.1155/2018/9293437

Kulkarni, A., Sthapit, A., Sedhain, A., Bhattarai, B., & Panthee, S. (2021). Texture Classification using Angular and Radial Bins in Transformed Domain. *International Journal of Advanced Computer Science and Applications*, *12*(3). https://doi.org/10.14569/IJACSA.2021.0120301

Mudinas, A., Zhang, D., & Levene, M. (2018). Bootstrap Domain-Specific Sentiment Classifiers from Unlabeled Corpora. *Transactions of the Association for Computational Linguistics*, *6*, 269–285. https://doi.org/10.1162/tacl_a_00020

Picozzi, M., & Iaccarino, A. G. (2021). Forecasting the Preparatory Phase of Induced Earthquakes by Recurrent Neural Network. *Forecasting*, *3*(1), 17–36. https://doi.org/10.3390/forecast3010002

Pilař, L., Kvasničková Stanislavská, L., & Kvasnička, R. (2021). Healthy Food on the Twitter Social Network: Vegan, Homemade, and Organic Food.

International Journal of Environmental Research and Public Health, *18*(7). https://doi.org/10.3390/ijerph18073815

Preto, A. J., Matos-Filipe, P., Mourão, J., & Moreira, I. S. (2022). SYNPRED: prediction of drug combination effects in cancer using different synergy metrics and ensemble learning. *GigaScience*, *11*, giac087. https://doi.org/10.1093/gigascience/giac087

Shafique, M. A., & Marchán, S. S. (2022). Investigating the Impact of Information Sharing in Human Activity Recognition. *Sensors*, *22*(6). https://doi.org/10.3390/s22062280

W. Li & B. Xu. (2020). Aspect-Based Fashion Recommendation With Attention Mechanism. *IEEE Access*, *8*, 141814–141823. https://doi.org/10.1109/ACCESS.2020.3013639

Wang, Z., & Wu, Q. (2018). An Integrated Deep Generative Model for Text Classification and Generation. *Mathematical Problems in Engineering*, *2018*, 7529286. https://doi.org/10.1155/2018/7529286

Wei, L., Song, Y., He, W., Chen, X., Ma, B., Lu, Y., & Zhu, X. (2020). Accuracy Improvement of IOL Power Prediction for Highly Myopic Eyes With an XGBoost Machine Learning-Based Calculator. *Frontiers in Medicine*, *7*. https://www.frontiersin.org/articles/10.3389/fmed.2020.592663