Time-series analysis of geographic depression scores on Twitter using BERT

Team 13

Final Report

AlOtaibi, Majed
majed@gatech.edu

Qabeel, Marwa F
mqabeel3@gatech.edu

Mugweru, Kevin W
mugweru@gatech.edu

Zamora Mennigke, Ricardo
rmennigke3@gatech.edu

Li, Jing
jli900@gatech.edu

Seedat, Hameeda
hsaif3@gatech.edu

1 INTRODUCTION

Recently, social media has become a part of many people's daily life. Social media are a set of digital platforms and sites used to exchange information in the form of text, audio, videos or photos and are accessed using different channels such as web and mobile phones. According to Sadagheyani and others, 4.5 billion people had internet connection in 2020 [16] and according to Naslund 3.8 billion of internet users are actually social media users [14], which represent around 84% of the total internet users. Due to the vast adoption of social media, people use different social media platforms of their choice to express their thoughts, ideas, emotions and feelings. This large amount of social media data gave researchers an opportunity to study and analyse users' emotions and feelings; specifically focusing on people's mental health and presenting the result to Healthcare organizations.

In this paper, we propose a depression detection model based on Twitter data with time series analysis to provide healthcare organizations with a visibility heat map of regional mental health trends. Our approach not only predicts depression with high accuracy but also offers additional context, such as regional depression trends and seasonality analysis, and visualization tools to guide healthcare and social workers in addressing the regions most impacted by mental health issues.

2 PROBLEM DEFINITION

In this digital era, Mental health is critical in everyone's life and the ability to analyze people's mental health and depression level is vital for Health organizations. Since many people use social media today for expressing their emotions and feelings, it's possible to utilize analytical methods and data visualization techniques to achieve the following objectives:

- Provide Health organizations with the required tools to get insights about people's mental health and depression level.
- Use time series analysis techniques to find patterns and seasonality about depression score overtime.
- Present findings using an easy to understand dashboard with different graphical representations such as heat map, wordcloud etc.

3 LITERATURE REVIEW

3.1 Twitter Data

Various AI algorithms or hybrid algorithms are used on twitter data for predicting depression. Khafaga [8] proposed a new approach called MDH-PWO for real-time detection of depression and anxiety disorders with high accuracy. Harnain [9] proposed a model that uses a combination of convolutional neural network (CNN) and long short-term memory (LSTM), with Word2Vec feature extraction technique.

The limitations from current approaches are: 1. Apart from giving depressed/not depressed binary prediction there is not enough context for mental health workers to provide better service. 2. There is no visualization available to guide on the regions impacted the most.

3.2 Bidirectional Encoder Representations from Transformers (BERT)

This project uses an NLP transformer called Bidirectional Encoder Representations from Transformers (BERT) [4], which is a pre-trained language model that has achieved better performance in sentiment classification techniques compared to other algorithms such as Logistic regression, LSTM [21] and SentiWordNet [1]. The model obtained an F1-score of more than 0.90 for depression detection [6].

Bidirectional Long Short Term Memory (BiLSTM) with Glove word embedding technique [18] [19] was used, whereby each word was represented with a vector of dimension 300; this got an F1 score of 0.65. As shown above, the BERT model has better accuracy than BiLSTM - on Reddit data, BERT-based classifiers were able to achieve a precision-score of 0.913 [13].

The most popular linguistic feature extraction method is word-embeddings [17]. However, embedding algorithms like word2vec and GloVe [11] lack the strength and understanding of natural language words in the context of the sentences they are used in [12]. Therefore, we will use the word-embedding internal to BERT.

Depressed user detection for a forum was conducted using clustering in a latent space of an auto-encoder with unsupervised learning[20]. This resulted in an F1 score of 0.64. This is less than the supervised approach taken by others [6] [18].

3.3 Time Series Analysis

Time Series Analysis is a technique used to analyse and extract information and patterns from data that is ordered chronologically over time. Time series analysis was used to identify critical time points and changes in sentiment [7]. Topic modeling was then applied to the tweets in the time series where significant sentiment shifts were identified. Time series analysis provided insights into customer behavior and identified areas where online retail brands needed to improve to enhance customer satisfaction. Bhullar and others [2] conclude that automation for time series analysis can optimize relief operation management to serve victims in the most efficient way. Moreover it can help in controlling legal and administration implications. This project uses time series analysis to analyse and identify depression trend and seasonality by country.

4 PROPOSED METHOD

Two value enhancing innovations are proposed:

1: Integrate time series analysis with Bert prediction results to provide healthcare organizations with more contextual information on the trend and seasonality of depression tweets by country.

2: Develop an interactive visualization tool enabling healthcare users to identify patterns of depression in the country of interest. By doing so, this tool will help healthcare organizations to plan and prepare their resources more effectively.

This project will use different analytical techniques to achieve analysis and create a depression detection model. The following sections will detail each technique along with the innovative ideas:

4.1 Twitter Data collection and cleanup

Code to scrape the data was obtained from a GitHub repository[1] that offers an automated scraper to extract Twitter data on a daily basis, we modified the code to obtain data from January 2020 to March 2023. The scraper targeted tweets using the hashtag "depression" and the data was stored in JSON files. The raw JSON dataset's size is approximately 3,5 GB. To extract relevant information from the dataset, we focused our analysis on several key features including "Tweet content," "Tweet date," "Tweet hashtags," "Retweet count," "Favorite count," and "Locations." We took care of 3 main features important to our analysis, "Tweet Content" by removing missing and duplicated tweets and other text transformations, mentioned in "Data Analysis" section. For the "Tweet hashtags", we removed some irrelevant hashtags from it, in addition to regular cleaning steps. Finally, we needed to get better location data, thus we used Google Maps API on "User Locations" to generate Country, City, Longitude, and Latitude data.

4.2 Data Analysis

Our framework follows the common text classification stages: pre-processing, tokenization, word embedding, model building, and training.

Pre-processing removes noise content within the text. Computers are not able to process text directly, which creates the need to transform text into quantitative values (tokenization) in the form of an array integer sequence. It is relevant to notice that a feature in each individual token occurrence frequency and the sample is the vector of all token frequencies for a given text.

[1] https://github.com/ahmedshahriar/depression-tweets-scraper

This way texts are described by word occurrences ignoring the position information of the words [15]. Word embedding is built into the BERT model. In language, quite often, there are words that are quite uninformative (such as "the") in depicting the content of a text. These are usually removed to improve model accuracy by preventing the model from identifying them as important features. [3].

In our approach, the BERT model is used. It is based on transformers which were pre-trained to find deep bidirectional representations of words in unlabeled text to later be used in language analysis tasks. The transformers are models that only use attention mechanisms that can generally be used to model or analyze sequences of objects that are not necessarily restricted to language [22].

The difference between BERT and the original transformer is that BERT is not autoregressive [5] and it only uses the encoder because its objective is to be a language model. Bert's speciality is its ability to process text in both directions simultaneously [22], that recurrent neural networks are not capable of. Moreover, BERT can relate 2 different sentences knowing that they do not belong to the same text, being very useful for question answering applications [4].

The model performance will be described through the accuracy. The final stage of the architecture is a fully connected layer with activation to output the approximation probabilities. The model is trained with a batch size of 64, an Adam optimizer, and 4 epochs.

4.3 Time Series Analysis

We can transform the labeled data obtained from the BERT model into time series data with three columns: tweet date, depressed scores and country. The depress scores are calculated by dividing the total count of depress label records of each Year Month by the total count of records (include positive and negative labels) of each Year Month, for each country.

To ensure sufficient data for time series analysis, we filtered out countries with less than 30 records. As a result, we were left with a total of 32 countries in our analysis. Figure 1 shows depressed scores by Year Month for United State as example.Using this data, we apply a 3-month moving average technique to smooth out fluctuations and noise by using the rolling function from the pandas library.

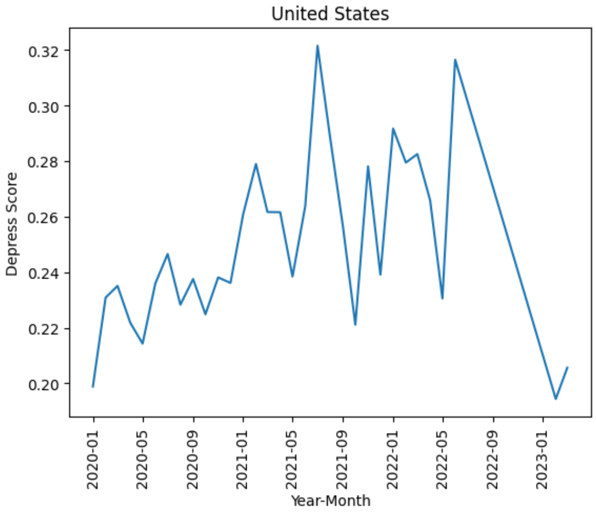


Figure 1: Monthly depressed scores.

We then use the seasonal decompose function with a period of 12 months to identify trends and patterns in the resulting moving average time series. The seasonal decomposition for United States is shown in Figure 2 as an example. We see that the depression score has been rising steadily month on month, with the depression score rate slowing a bit after the first 2 months (approximately) of 2021. We also see that depression score tends to peak around or just after the mid year. Mental health organisations should then use this information to craft a strategy around preventing depression rates in the US. The same is true for other countries within the data.

4.4 Presentation and Visualization

Tableau is used to create a story that explores 3 interactive dashboards: The first dashboard is a general data exploration for the most important features for our analysis like hashtags, number of tweets, and "tweets favorite" count. A word cloud of hashtags from depressed tweets was plotted, and a map chart that shows countries and their subsequent cities was used to show the total tweets count. Also, there is a box-plot distribution of tweet counts and replies, the plot includes the five-number summary of the plotted data against the investigated years in the dataset.

The second dashboard shown in "Figure 3" displays the time series analysis results. It shows a calculated depression score over time per country using a map view,

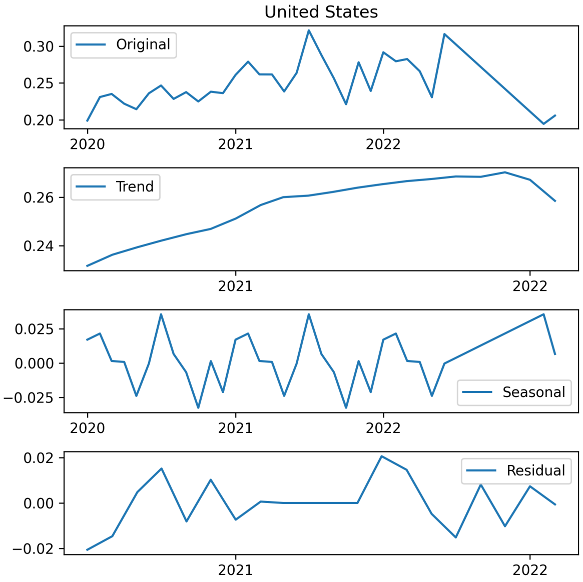
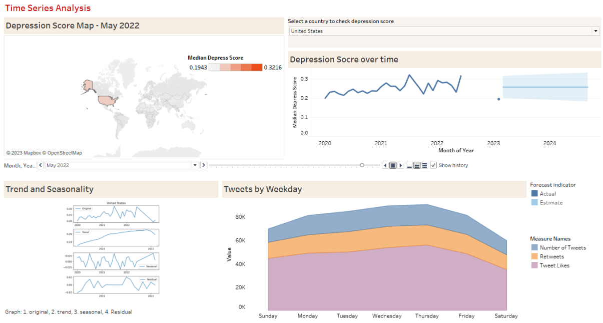


Figure 2: US Seasonal decomposition


Time Series Analysis

Depression Score Map - May 2022

Select a country to check depression score: United States

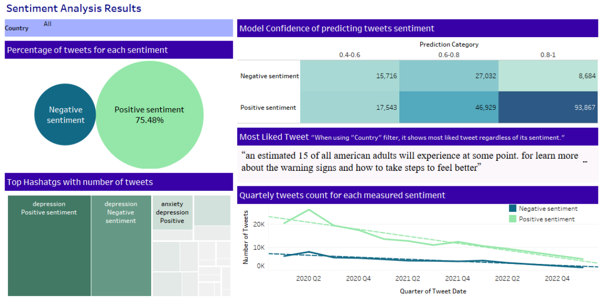
Depression Socre over time

Trend and Seasonality

Tweets by Weekday

Graph: 1. original, 2. trend, 3. seasonal, 4. Residual

Figure 3: Time Series Analysis dashboard


Sentiment Analysis Results

Country: All

Percentage of tweets for each sentiment

Model Confidence of predicting tweets sentiment

	Prediction Category		
	0.4-0.6	0.6-0.8	0.8-1
Negative sentiment	15,716	27,032	8,684
Positive sentiment	17,543	46,929	93,867

Most Liked Tweet "When using "Country" filter, it shows most liked tweet regardless of its sentiment."

"an estimated 15 of all american adults will experience at some point. for learn more about the warning signs and how to take steps to feel better"

Top Hashatgs with number of tweets

Quartely tweets count for each measured sentiment

Figure 4: Sentiment Analysis dashboard

a line chart, an area chart, and decomposed graphs. The map view includes a Pages card that allows users to control and observe changes in depression scores over time. The line chart has a forecast function to project potential future depression scores. The decomposed graphs, from the time series analysis, display the trend and seasonal patterns per country. Users can filter for specific countries of interest. The area chart shows the volume of depressed tweets by day of the week, allowing us to identify which day of the week has the highest frequency of depressed tweets.

The third dashboard shows results from our Bert model prediction. A bubble chart is used to show the percentage of total tweets for each sentiment; "Negative" and "Positive", this chart is being used as a filter to check the changes in the other elements of the dashboard. A Highlight table is used to show the distribution of the model predictions probability or model confidence in detecting the tweets as "Positive" or "Negative" sentiments across 3 categories created using a calculated field in Tableau. There is also a treemap chart created to show the top hashtags based on the number of tweets and which of them appeared in a "Positive" or "Negative" context. A quarterly line graph is drawn to show the trend of tweets for each sentiment group. Lastly, there is a text mark that displays the top liked tweet regardless of the tweet sentiment, these tweets work dynamically with "Country" filter placed at the top of the dashboard, when changing the country, the text mark will show the top liked tweet in this country. "Figure 4" shows "Dashboard 3" in the Story which presents the sentiment analysis results. Dynamic filters are used to increase the interactivity in each dashboard. Also, the user can utilize tooltips on any chart by hovering over the chart. There are no Null data in the dataset, they all have been taken care of in the pre-processing.

5 EXPERIMENTS AND EVALUATION

5.1 Inquiries that our experiments aim to address

BERT-trained model, time series analysis technique, and visualizations were designed to provide healthcare workers, or anyone interested in this study, with answers to the following questions:

Bert trained model would provide:

• How many tweets associated with "#depression" were classified as negative or "depressed" vs. positive or "non-depressed" over the time period of the gathered data?

- What is the model's accuracy in detecting depression?
- What was the model's confidence in determining the sentiment of tweets as "Negative" or "Positive"?

The time series analysis results would answer:

- How have depressed scores changed over time in different countries?
- Are there any seasonal patterns or trends in the depressed scores per country?

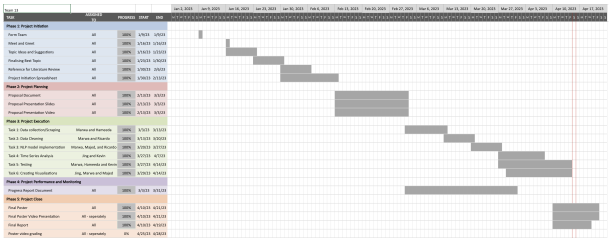
Data visualizations would show:

- Were there any other hashtags that were significantly associated with "#depression" based on tweet count?
- What is the percentage of tweets marked as "Negative" or "Positive" from the country/city level?
- What are the distributions of model confidence categories based on the number of tweets gathered for each sentiment group?

5.2 Detailed description of the Experiments and Evaluation

- Experiment 1 scraped Twitter for tweets with the "#happiness" hashtag, we removed any positive tweets that were mixed with "#depression" dataset and did the same for "#happiness" dataset. We then randomly selected an equal number of tweets from both datasets and trained our model. This resulted in poor performance during training and testing.
- Experiment 2 included all tweets with the "#depression" hashtag, regardless of positive or negative connotations. We only excluded tweets with irrelevant keywords such as countries, states, or cities. This approach allowed our sentiment model to determine whether a tweet has a positive or negative sentiment in the context of the emotion being measured; "depression". Also, we used English tweets only to avoid confusing the sentiment model. And because using "googletrans" Python library on non-English tweets didn't return good results. This experiment was more successful than the first, hence we decided to use this approach.

After we finished data pre-processing, we excluded part of the dataset with only "3200" tweets, we manually labeled the tweets of this subset as "Positive" or "Negative". Then we used this subset to train the model, the data splitted into 80% for training and 20% for testing.



Team 13

TASK	ASSIGNED TO	PROGRESS	START	END
Phase 1: Project Initiation				
Form Team	All	100%	1/9/23	1/9/23
Meet and Greet	All	100%	1/16/23	1/16/23
Topic Ideas and Suggestions	All	100%	1/16/23	1/23/23
Finalising Best Topic	All	100%	1/23/23	1/30/23
Reference for Literature Review	All	100%	1/30/23	2/6/23
Project Initiation Spreadsheet	All	100%	1/30/23	2/13/23
Phase 2: Project Planning				
Proposal Document	All	100%	2/13/23	3/3/23
Proposal Presentation Slides	All	100%	2/13/23	3/3/23
Proposal Presentation Video	All	100%	2/13/23	3/3/23
Phase 3: Project Execution				
Task 1: Data collection/Scraping	Marwa and Hameeda	100%	3/3/23	3/13/23
Task 2: Data Cleaning	Marwa and Ricardo	100%	3/13/23	3/20/23
Task 3: NLP model implementation	Marwa, Majed, and Ricardo	100%	3/20/23	3/27/23
Task 4: Time Series Analysis	Jing and Kevin	100%	3/27/23	4/7/23
Task 5: Testing	Marwa, Hameeda and Kevin	100%	3/27/23	4/14/23
Task 6: Creating Visualisations	Jing, Marwa and Majed	100%	3/29/23	4/14/23
Phase 4: Project Performance and Monitoring				
Progress Report Document	All	100%	3/3/23	3/31/23
Phase 5: Project Close				
Final Poster	All	100%	4/10/23	4/21/23
Final Poster Video Presentation	All - seperately	100%	4/10/23	4/21/23
Final Report	All	100%	4/10/23	4/19/23
Poster video grading	All - seperately	0%	4/25/23	4/28/23

Figure 5: Gannt chart showing high level plan.

Further, to ensure the model's predictive accuracy was not induced by overfitting:

i. We monitored model performance on both the training and testing sets over time using "TensorBoard" feature in PyTorch.

ii. The pre-trained model, "BertForSequenceClassification" has a dropout of 0.1 by default; this helps prevent overfitting by randomly dropping out some neurons during training. It prevents the model from relying too heavily on specific instances in the training set.

We would like to shed light on the fact that we worked on adjusting the model by splitting the data to 80:10:10 for training, validation, and testing respectively. However, finding the right combination to tune the model hyper-parameters, training, validating then testing the data was time consuming; even using GridSearch technique. Hence, we decided to not proceed working with the new model and instead stuck to our original model explained in this section. Our intention was to use cross-validation, which involves splitting the data into multiple subsets (folds), training the model on each subset while validating it on the remaining data, and then averaging the results to get an estimate of the model's performance on new, unseen data.

6 CONCLUSION AND DISCUSSIONS

The project spanned over five months. This includes project initiation, planning, execution, and closing. Tasks were equally distributed amongst team members as summarized in Figure 5.
In this project we used an NLP transformer called (BERT) to analyze scraped twitter data to achieve depression

score analysis. The model was trained on subset of the data and achieved a high accuracy of 80%. BERT was fine-tuned using a training set and evaluated on a testing set using metrics like accuracy, precision, recall, and F1-score, see Table 1. The model has a high accuracy of 0.80, but the precision for negative classification is low at 0.58. However, the precision for positive classification is high at 0.88, and the recall values for both classes are relatively high, indicating that the model is able to identify positive and negative instances to a large extent. The F1-score values for both classes are also relatively high, although lower for positive classification. Additionally, we used a confusion matrix (check "Figure 6") to evaluate the model performance. True positives with (63.75%) was the result of correctly predicting positive cases, while true negatives with (16.25%) represent the percentage of correctly predicted negative cases. False positives with an (8.12%): the result of negative cases that were incorrectly predicted as positive, and false negatives with (11.88%) were positive cases that were incorrectly predicted as negative. Based on this result, the model succeeded to achieve high accuracy with a higher proportion for the true positives and true negatives.

We applied the trained and tested model to make predictions on completely new data. The subset of data that was excluded in the beginning and labeled manually was not included in this new dataset that was used for predictions, to ensure that there was no bias. As mentioned before, training and validating the data would be more appropriate, it would help detect overfitting and ensure the model isn't making inaccurate generalizations on the new data. However, due to this model requiring high GPU which was limited for our team, and the time required to tune the hyperparameters; we had to proceed with the same model to make predictions. We also wanted to label more data to use for training the model, we have only 3200 labelled rows which is relatively small. We would like to proceed with more experiments in the future, we reached 72% accuracy with the new model, and the availability of more time and more labeled data would help us produce a more robust model.

After getting sentiment results from the model, we conducted a time series analysis to display trends and seasonality using the proposed depression score for each country. Our analysis reveals interesting insights: different countries have different trends of depression scores from 2020 to 2023. However, there is a common seasonal trend across most countries; that is, depressed tweets occur more frequently during the winter. This is aligned with studies that suggest seasonal affective disorder (SAD) is caused by a lack of vitamin D3 during the winter periods [10].

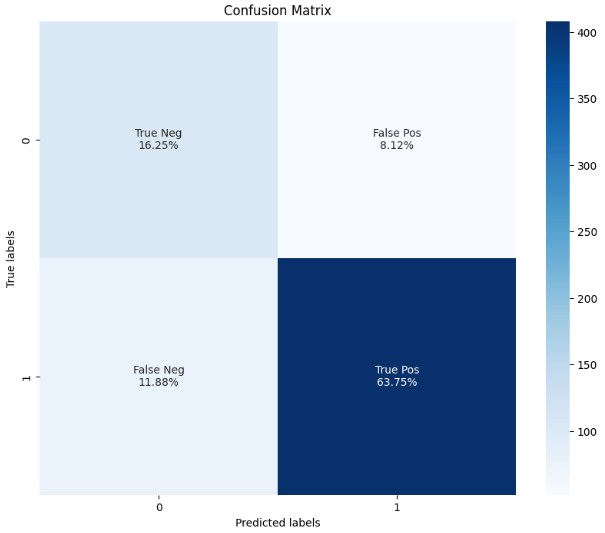


Figure 6: Confusion Matrix

For data visualization, Tableau was used to build an interactive dashboard that explores 3 sections; general data exploration, time series analysis, and Bert model prediction results.

The approach used in the report is very promising. However, due to insufficient data for some countries, we only included around 30 countries in the dashboard for time series analysis. It is important to note that many aspects of this approach are based on project team assumptions, and these assumptions require validation from a health organization for evaluation and potential adjustments if needed.

Table 1: Model Evaluation Results

	precision	recall	f1-score
0	0.58	0.67	0.62
1	0.89	0.84	0.86
accuracy	0.8	0.8	0.8
macro avg	0.73	0.75	0.74
weighted avg	0.81	0.80	0.80

REFERENCES

[1] Shivaji Alaparthi and Manit Mishra. 2021. BERT: a sentiment analysis odyssey. *J. Mark. Anal.* 9, 2 (June 2021), 118–126. https://doi.org/10.48550/arXiv.2007.01127

[2] Gurman Bhullar, Aseem Khullar, Apoorva Kumar, Anirudh Sharma, H S Pannu, and Avleen Malhi. 2022. Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient resource management. *Int. J. Disaster Risk Reduct.* 75, 102979 (June 2022). https://doi.org/10.1016/j.ijdrr.2022.102979

[3] Francois Chollet et al. 2015. *Keras.* https://github.com/fchollet/keras

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL]

[6] Mohammad El-Ramly, Hager Abu-Elyazid, Youseef Mo'men, Gameel Alshaer, Nardine Adib, Kareem Alaa Eldeen, and Mariam El-Shazly. 2021. CairoDep: Detecting Depression in Arabic Posts Using BERT Transformers. In *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*. 207–212. https://doi.org/10.1109/ICICIS52592.2021.9694178

[7] Noor Farizah Ibrahim and Xiaojun Wang. 2019. Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media. *Computers in Human Behavior* 96 (10 2019). https://doi.org/10.1016/j.chb.2019.02.004

[8] Doaa Sami Khafaga, Maheshwari Auvdaiappan, K Deepa, Mohamed Abouhawwash, and Faten Khalid Karim. 2023. Deep Learning for Depression Detection Using Twitter Data. *Intelligent Automation & Soft Computing* 36, 2 (2023). https://doi.org/10.18653/v1/W18-0609

[9] Harnain Kour and Manoj Kumar Gupta. 2022. Predicting the language of depression from multivariate twitter data using a feature-rich hybrid deep learning model. *Concurrency and Computation: Practice and Experience* 34, 24 (2022). https://doi.org/10.1002/cpe.7224

[10] Allen TG Lansdowne and Stephen C Provost. 1998. Vitamin D3 enhances mood in healthy subjects during winter. *Psychopharmacology* 135 (1998), 319–323. https://doi.org/10.1007/s002130050517

[11] Paula Lopez-Otero, Laura Docío Fernández, Alberto Abad, and Carmen Garcia-Mateo. 2017. Depression Detection Using Automatic Transcriptions of De-Identified Speech. https://www.isca-speech.org/archive_v0/Interspeech_2017/pdfs/1201.PDF.

[12] Ben Lutkevich. 2020. What is Bert (language model) and how does it work? https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model.

[13] Rodrigo Martínez-Castaño, Amal Htait, Leif Azzopardi, and Yashar Moshfeghi. 2021. BERT-Based Transformers for Early Detection of Mental Health Illnesses. In *Lecture Notes in Computer Science*. Springer International Publishing, 189–200. https://doi.org/10.1007/978-3-030-85251-1_15

[14] John A. Naslund, Ameya Bondre, John Torous, and Kelly A. Aschbrenner. 2020. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. *Journal of Technology in Behavioral Science* 5, 3 (4 2020). https://doi.org/10.1007/s41347-020-00134-x

[15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830. https://doi.org/arXiv:1201.0490

[16] Hassan Ebrahimpour Sadagheyani and Farin Tatari. 2020. Investigating the role of social media on mental health. *Mental Health and Social Inclusion* 25 (10 2020). https://doi.org/10.1108/mhsi-06-2020-0039

[17] Rafael Salas-Zárate, Giner Alor-Hernández, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde, Maritza Bustos-López, and José Luis Sánchez-Cervantes. 2022. Detecting Depression Signs on Social Media: A Systematic Literature Review. *Healthcare* 10, 2 (2022). https://doi.org/10.3390/healthcare10020291

[18] Faisal Muhammad Shah, Farzad Ahmed, Sajib Kumar Saha Joy, Sifat Ahmed, Samir Sadek, Rimon Shil, and Md. Hasanul Kabir. 2020. Early Depression Detection from Social Network Using Deep Learning Techniques. In *2020 IEEE Region 10 Symposium (TENSYMP)*. 823–826. https://doi.org/10.1109/TENSYMP50017.2020.9231008

[19] Ying Shen, Huiyu Yang, and Lin Lin. 2022. Automatic Depression Detection: an Emotional Audio-Textual Corpus and A Gru/Bilstm-Based Model. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 6247–6251. https://doi.org/10.1109/ICASSP43922.2022.9746569

[20] Anu Shrestha, Edoardo Serra, and Francesca Spezzano. 2020. Multi-modal social and psycho-linguistic embedding via recurrent neural networks to identify depressed users in online forums. *Network Modeling Analysis in Health Informatics and Bioinformatics* 9 (12 2020). https://doi.org/10.1007/s13721-020-0226-0

[21] Abdul Hasib Uddin, Durjoy Bapery, and Abu Shamim Mohammad Arif. 2019. Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique. In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. 1–4. https://doi.org/10.1109/IC4ME247184.2019.9036528

[22] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. arXiv:1706.03762 [cs.CL]