**AI Certificate- Capstone Project**

**Project Report**

**Detecting Depression in Arabic Tweets**

Ahmad Hamad

May 2023

# **Introduction**

A mental disorder is characterized by a clinically significant disturbance in an individual’s cognition, emotional regulation, or behavior [1]. One of the most common mental disorders is depression. Depression is characterized by prolonged and inescapable low mood, often accompanied by a loss of interest in previously enjoyable things. Depression in the Arab world is on the rise, with approximately three in 10 people across the region saying they suffer from depression [2].

Due to many reasons, the Arabic region faces many external triggers, which could intensify mental health disorders such as widely experienced trauma due to war, violent regimes, and displacement. In addition, there are other reasons behind increasing depression in the Arabic region such as Lack of mental health awareness and psychological help, Increase usage of social media applications and less social interaction, and Lack of physical activity [3].

Nowadays, social media platforms are becoming where a person can express their feelings and opinions and show reactions using texts, videos, or images. One of the most common of these social media platforms in the Arabic world is Twitter where people can express their feelings and emotions openly through texts called Tweets. Many of these tweets can help detect if the writer has depression. This project aims to build a machine-learning model to detect Arabic tweets if they include content with depression tendencies or not, which can help in the earlier detection of depression.

# **Problem Statement**

Depression is one of the most common mental disorders in the Arab region, and many Arabs do not consider depression a real disease that can affect their lives. This project proposes a methodology for detecting depression in Arabic tweets to classify if they contain content with tendencies to depression or not, which helps in the early detection of depression and share awareness among the Arab community about depression as a real illness that threats a person's life.

**Previous Work**

Depression is a prevalent mental health disorder that affects individuals worldwide. Social media platforms have become a popular medium for individuals to express their emotions and share their experiences. Researchers have explored the possibility of detecting depression in social media data using various strategies and models. In this literature review, we summarize three studies that aim to detect depression in social media data.

M. R. Islam et al. [4] used crowdsourcing and data collected from public social media accounts to detect depression. The study employed four machine learning algorithms, namely Random Forest, Naive Bayes, AdaBoostM1, and Liblinear, to classify tweets as either depression or non-depression. The Liblinear algorithm achieved the highest accuracy and recall in detecting depression, which was attributed to its ability to perform efficiently in large-scale linear classification with better performance in training time and precision. While S. Almouzini and A. Alageel [5] compared two deep learning models, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), with Support Vector Machine (SVM) for depression detection from Arabic social media data. The study found that CNN outperformed the other two models in detecting depression, achieving an accuracy of 88% and a recall of 82%. The study concluded that CNN is a better solution for depression detection from Arabic social media data. Moreover, M. El-Ramly et al. [6] employed pre-trained language models to detect depression in Arabic dialectic tweets. The study used eleven pre-trained language models, including BERT, GigaBERT, XLM-RoBERTa, AraBERT, AraELECTRA, Arabic BERT, Arabic ALBERT, ARBERT, MARBERT, QARiB, and AraGPT2. The study found that MARBERT was the best-performing model for detecting depression in Arabic dialectic tweets, achieving a macro-average F1-score of 88.75%.

Overall, these studies provide insights into the different approaches that can be used to detect depression in social media data and the performance of various models in this task. However, further research is needed to explore the generalizability of these models to other languages and cultures and to address ethical concerns related to privacy and data protection.

**Final proposed approach**

In order to detect depression in Arabic tweets, a deep learning model was built using AraBERTv0.2-Twitter-base [7], a variant of BERT (Bidirectional Encoder Representations from Transformers), which is a pre-trained transformer-based language model for natural language processing (NLP) tasks. It was introduced by Google in 2018 and has since then achieved state-of-the-art performance on a wide range of NLP tasks, such as sentiment analysis, question answering, and text classification.

The architecture of BERT is based on the Transformer, which is a type of neural network architecture that was introduced in a 2017 paper titled "Attention Is All You Need" by Vaswani et al. Transformers are different from traditional neural networks in that they can process entire input sequences at once, rather than processing them sequentially or in parallel. This means they can capture long-range dependencies and relationships between words in a sentence more effectively than traditional neural networks.

The key component of the Transformer architecture is the attention mechanism, which allows the model to weigh the importance of different parts of the input sequence when making predictions. Therefore, AraBERTv0.2-Twitter-base has been specifically trained on Arabic Twitter data for sentiment analysis tasks. The model is fine-tuned on a dataset of labeled Arabic tweets, with the goal of predicting whether a given tweet is indicative of depression or not.

 AraBERTv0.2-Twitter-base model is chosen for this task because it has been pre-trained on a large corpus of Arabic text and fine-tuned specifically on Arabic Twitter data for sentiment analysis tasks, making it an effective choice for detecting depression in Arabic tweets.

*Experimental Setup:*

The project's experimental setup involves six steps, which are:

1. Data Collecting: The data for this study were collected from two sources. The first dataset was obtained from the paper "[AraDepSu: Detecting Depression and Suicidal Ideation in Arabic Tweets Using Transformers](https://aclanthology.org/2022.wanlp-1.28.pdf)" [8], which consists of 20214 rows labeled into three classes:

1) Non-depression (12574 rows).

 2) Depression Mood (5472 rows).

 3) Depression with Suicidal Ideation (2167 rows).

The second dataset was obtained from the GitHub repository "Depression-detector"[9], which consists of 1058 rows labeled into two classes:

1) Depressed (529 rows).

2) Not depressed (529 rows).

2. Data preprocessing:

1. Datasets Integrating

Before integrating the above two datasets there are few steps:

* Merging the two labels “Depression Mood” and ”Depression with Suicidal Ideation” to be one label “Depression Mood” as the goal of this project is just detecting depression and not suicidal ideation because depression usually is linked to a greater risk of suicide.
* Merging these two datasets and changing the labels to “1”= depression and “0” = no depression.

2) Dropping duplicate rows.

        3) Dropping Nan rows.

        4) Handling imbalanced data:

The distribution of classes in the dataset is unequal, with the depression class having significantly fewer examples than the other (no depression=13088 instances while depression= 8146 instances). To handle the unbalancing, the majority class was undersampling by removing some examples using the sklearn resample[10]. Figures 4 and 5 show the data before and after undersampling.

A blue and orange pie chart

Description automatically generated

*Figure 1. Data before undersampling*

A picture containing text, screenshot, circle, diagram

Description automatically generated

*Figure 2. Data after undersampling*

5) The Arabert model's preprocessing function was applied to the dataset, which involved several operations, such as removing HTML markup, replacing URLs, emails, and mentions with special tokens, removing diacritics and tatweel, replacing repetition of more than two non-digit characters with two of the same character, mapping Hindi numbers to Arabic and applying Farasa segmentation.

3. Model fine-tuning:

Hyperparameter tuning is an essential step in fine-tuning an Arabert transformer model to achieve the best possible performance on a specific natural language processing (NLP) task. Hyperparameters are configuration settings that affect how the model is trained and how it learns from the task-specific dataset. Tuning these hyperparameters involves selecting the optimal values for them to maximize the model's performance on the target task.

These hyperparameters are:

* learning\_rate: 2e-05
* train\_batch\_size: 16
* seed: 25
* optimizer: Adam with betas=(0.9,0.999) and epsilon=1e-08
* lr\_scheduler\_type: linear
* num\_epochs: 8

4. Evaluation: We found that Bert and Bleu score are commonly used metrics in natural language processing (NLP) tasks, including sentiment analysis and language translation. However, these metrics may not be suitable for evaluating the performance of the Arabert model in detecting depression in Arabic tweets.

Bert and Bleu score focus on evaluating the quality of the generated text or translation, but they do not take into account the specific task at hand. In the case of detecting depression in Arabic tweets, accuracy is a more appropriate metric since it measures the percentage of correctly classified tweets, which is directly related to the task's objective.

Furthermore, the Arabert model's output is binary (depressed or not depressed), which makes the classification task simpler than generating text or translations. Therefore, evaluate the performance of the fine-tuned model using various metrics, such as accuracy, precision, recall, F1-score, and kappa. This step helps to ensure that the model generalizes well to unseen data.

5. Deployment: the model will be deployed using Gradio and HuggingFace, which provide the framework for the fast building of deep learning apps.

**Results, Discussion, and Outcomes**

The proposed model showed high performance in all the evaluation metrics including accuracy, precision, recall, F1-score, and kappa. Table 1 shows the results.

*Table 1. Results*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Precision** | **Recall** | **Accuracy** | **F1-score** | **Kappa** |
| 0.9281 | 0.9279 | 0.9279 | 0.9279 | 0.8558 |

From the table above we can see that the model achieved an accuracy of 93.97% with the same F1-score and recall, while for the precision the model’s performance was a bit higher with a performance of 92.81%. The kappa metric shows different results with 85.58%.

The model’s performance could be affected by many factors. First, performing more hyperparameter tuning could improve the performance of the model. For example, using more epochs for training the model can result in significant improvement. Moreover, using a larger dataset could result in better performance, or maybe using different resampling techniques such as oversampling.

In addition, the model after deployment showed accurate results and proved its performance in real-life usage. Figure 3 shows the deployed model.

A screenshot of a web page

Description automatically generated with medium confidence

*Figure 3. The deployed model*

**Implications of The Results**

The proposed model is an efficient solution to detect depression early and easily. As the model has been deployed and is publicly available can be used by companies to tech the mental health of their employees. Also, can help therapists to check the tweets of their patients to analyze their thoughts. Moreover, the fact that the model is easy for using will help people to check on their own health before making the decision of starting therapy. Many use cases are possible of the proposed models and many benefits can be gained, and this will be a jump in the mental health field and hopefully will reduce the side effects of depression.

**Potential Directions for Future Work**

To detect depression in Arabic tweets, it is crucial to have access to up-to-date and relevant data. To achieve this, an ETL (Extract, Transform, Load) pipeline can be built to stream Twitter data in real time. This pipeline will ensure that the data used for the detection model is constantly updated and relevant. Furthermore, integrating multimodal data, such as text, images, and videos, into the detection model can provide a more comprehensive and accurate understanding of a user's mental health status. This can be achieved by developing a machine-learning model that can analyze both textual and visual data from tweets. Finally, to take the next step in helping individuals who have been identified as depressed, a chatbot can be developed that uses the data gathered by the detection model to provide personalized recommendations and resources for the individual. This chatbot can also help identify potential causes of depression by asking targeted questions and providing feedback based on the individual's responses.

# References

[1] “Mental disorders.” https://www.who.int/news-room/fact-sheets/detail/mental-disorders (accessed Apr. 15, 2023).

[2] Addcounsel, “Depression in Arabic Countries,” Addcounsel, Apr. 13, 2022. https://addcounsel.com/depression-in-arabic-countries/ (accessed Apr. 15, 2023).

[3] D. H. Naveed, “Mental Health in the Middle East,” Middle East Medical Portal, Sep. 10, 2018. https://www.middleeastmedicalportal.com/mental-health-in-the-middle-east/ (accessed Apr. 15, 2023).

[4] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, “Depression detection from social network data using machine learning techniques,” Heal. Inf. Sci. Syst., vol. 6, no. 1, pp. 1–12, Dec. 2018, doi: 10.1007/s13755-018-0046-0.

[5] S. Almouzini and A. Alageel, “2 Salma Almouzini, Maher khemakhem,” Asem Alageel/ Procedia Comput. Sci., vol. 163, pp. 0–000, 2019, doi: 10.1016/j.procs.2019.12.107.

[6] M. El-Ramly et al., “CairoDep: Detecting Depression in Arabic Posts Using BERT Transformers,” Proc. - 2021 IEEE 10th Int. Conf. Intell. Comput. Inf. Syst. ICICIS 2021, pp. 207–212, 2021, doi: 10.1109/ICICIS52592.2021.9694178.

[7] “aubmindlab/bert-base-arabertv02-twitter · Hugging Face,” Apr. 05, 2023. https://huggingface.co/aubmindlab/bert-base-arabertv02-twitter (accessed May 06, 2023).

[8] M. Hassib, N. Hossam, J. Sameh, and M. Torki, “AraDepSu: Detecting Depression and Suicidal Ideation in Arabic Tweets Using Transformers,” in Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP), Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, Dec. 2022, pp. 302–311. [Online]. Available: https://aclanthology.org/2022.wanlp-1.28

[9] alhuri, Depression-detector. 2023. Accessed: May 06, 2023. [Online]. Available: https://github.com/alhuri/Depression-detector

[10] “sklearn.utils.resample,” scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.utils.resample.html (accessed May 06, 2023).