

Identifying Level of Depression with the Help of Machine Learning

Abdullah Al Rifat	Tanvir Ahmed
2016-2-60-030	2016-1-60-010
alrifatabdullah@gmail.com	tanvir.rafee405@gmail.com

F M Shahriar Ferdous
2016-2-60-012
fsyshahriar127@gmail.com

September 26, 2020

1 Introduction

Depression is the most common psychological disorder that creates strenuous complexities while treating with existing medications. According to WHO, more than 264 million people are affected worldwide.[1] This is a serious condition that can affect anyone from any age group. The severity of this malady enforced in advancement of various machine learning methods to assist the treatment process of depressive disorders.

Neuroscientists around the world are utilizing machine learning to produce steady treatment for patients.[2] By implementing Artificial Neural Network (ANN) the model will be able to provide a rational and logical outcome to predict if a person is depressed or not. We have obtained the raw dataset for this project from a reliable source. The following step is to apply various data engineering methods to process the data and divide it into two parts, test set(0.2), and train set. So that we can train the ANN on eight thousand observations and test its performance on two thousand observations. Moreover, the goal is to build a classifier.

1.1 Objectives

- To prognosticate the depression level among adult demographic.
- Several groups will be identified according to the severity of depression based on our model's prediction.
- An insight into the tremendous hardship that people are going through due to depression and propose an intuitive solution to minimize the damage.

1.2 Motivation

Millions of people are experiencing depression but still, it hasn't been taken seriously in many countries. In many places, there are not even adequate facilities to have an actual diagnosis. Besides, the lack of proper treatment and diagnosis makes the patient's condition more detrimental over time. Many get overwhelmed by their situation that it reaches to a boiling point, which results in self-destruction.

If our model can accurately predict the actual level of someone being depressed that will be unprecedented. Even though there are already existing machine learning models showing significant promise in this delicate sector. But still, there are tremendous possibilities to improve on a large scale with more research.

1.3 Existing Works

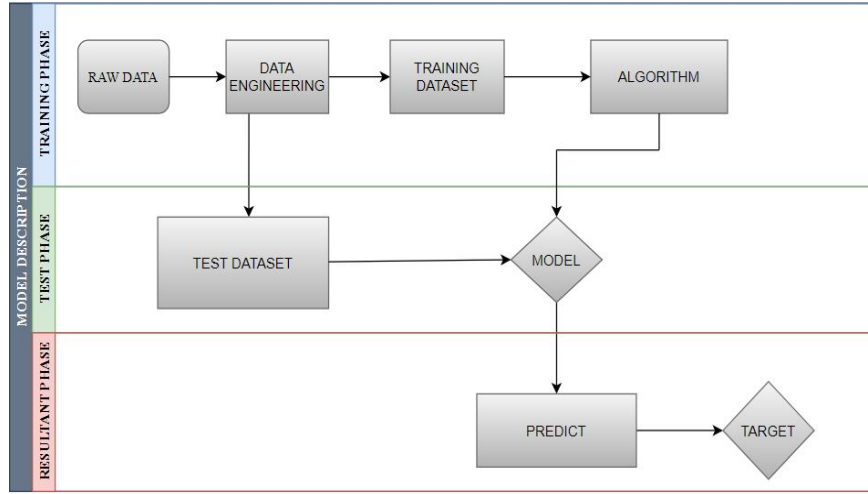
Researchers have been working on depression analysis for many years. Mostly in developed countries, there are various types of research still under development phase as it carries immense significance. Implementing Deep Convolutional Neural Network (DCNN) to detect depression from speech has been proposed already along with several other research works.[3] A survey paper on existing depression detection systems has published that incorporated the most competent practices in this delicate sector.[4]

1.4 Necessity

The core purpose of this project is to disentangle a multifaceted problem that is living and breathing in the heart of every modern-day society. The model we have developed over time is capable of predicting the level of depression with high accuracy, which can easily catch the attention of ambitious researchers both in the medical and research field. Nonetheless, it can minimize the hassle of medical professionals in detecting specific cases to a certain extent.

2 Methodology

We have collected the dataset from a well-known source. After that, we had to omit a few features from the dataset and split it into test and train set to develop the model with Artificial Neural Network. The model is capable of predicting the depression level for a particular case with adequate outcomes.



3 Implementation

3.1 Data Collection

The dataset is based on the people living in rural areas. All the columns were not clarified properly so the dataset was preprocessed to remove some features. The final features were survey id, village id, sex, age, marriage status, number of children, education level, total family members, gained asset, durable asset, saved asset, living expenses, other expenses, salary, incomes from own farm, business, no business, agriculture, farm expenses, labor primary, lasting investment, non-lasting investment, depressed: [Zero: No depressed] or [One: depressed].

3.2 Data Processing

There were no notable preprocessing methods applied but, we have omitted two features (survey id, village id) from the existing dataset so that any irrelevant data does not affect during the training phase of the model.

3.3 Model Development

We used Artificial Neural Network (ANN) for creating the model. The following neural network libraries 'Keras' and 'Tanserflow' were used in the backend of Artificial Neural Network(ANN). We split the dataset into train set and test set and then called the Keras library and imported some functions for developing the ANN model. There are 100 hidden layers and batch size is 10. For the input, we have used Rectified Linear Unit (ReLU) and for output the Sigmoid function. From the model, we reached an accuracy of 85.91

3.4 Results

We have got these results so far after implementing the model. (TP; the model accurately identified as depressed), (TN; the model accurately identified as non-depressed), (FP; non-depressed cases were identified as depressed), and (FN; depressed cases were identified as non-depressed).

$$TruePositive = 229$$

$$TrueNegative = 7$$

$$FalsePositive = 12$$

$$FalseNegative = 38$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 0.834$$

$$Precision = \frac{TP}{TP + FP} = 0.98$$

$$Recall = \frac{TP}{TP + FN} = 0.84$$

$$Specificity = \frac{TN}{TN + FP} = 0.25$$

$$F1Score = .90$$

4 Conclusion

In this project, we have developed a machine learning model with the help of Artificial Neural Network (ANN) which can predict the depression level from any classifier based dataset. Major mental illness like depression can also be detected and solved by integrating machine learning.

4.1 Challenges

Selecting the appropriate machine learning model was the main challenge to complete this project. Along with that, finding the relevant dataset was somewhat of a challenge within the short period of time. Setting up the appropriate environment was another challenge to develop the model as many libraries get deprecated over time.

4.2 Limitations

Our model is very useful for solving classifier-based problems but in terms of the categorical dataset, it may not perform up to the expectations. In that case, we may have to preprocess the data before training the model to achieve more accuracy.

4.3 Future Direction

Our model can perform well on any type of real-life classifier dataset based problems, for instance a customer choosing the right insurance policy or bank loan, purchasing decisions of a luxury product, and in many disease detection. The accuracy of our model can be improved by numerous training with unambiguous dataset and more research.

References

- [1] WHO, “Depression.” <https://www.who.int/news-room/fact-sheets/detail/depression>, Jan. 2020.
- [2] A. Z. M. S. A. R. . C. R. Victor, E., “Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation.,” *Psychological Assessment*, vol. 31, no. 8, p. 1019–1027, 2019.
- [3] C. LangHea, “Automated depression analysis using convolutional neural networks from speech.,” *Journal of Biomedical Informatics*, vol. 83, pp. 103–111, 2018.
- [4] R. L. Michelle Morales, Stefan Scherer, “A cross-modal review of indicators for depression detection systems.,” vol. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality, pp. 1–12, 2017.