
PREDICTION AND DETECTION OF DEPRESSION USING FOREST TREE ALGORITHM

ABSTRACT

Depression is a common illness that negatively affects feelings, thoughts and behaviors and can harm regular activities like sleeping. It is a leading cause of disability and many other diseases (Choudhury, et al 2013, Mathur et al, 2016, Watkins et al, 2013). According to WHO (World Health Organization) 1 statistics, more than 300 million people over the world are affected in depression and in each country at least 10% are provided treatment. Poor recognition and treatment of depression may aggravate heart failure symptoms, precipitate functional decline, disrupt social and occupational functioning, and lead to an increased risk of mortality (Cully, et al 2009). Early detection of depression is thus necessary. Unfortunately the rates of detecting and treating depression among those with medical illness are quite low (Egede, 2007). This research proposes a solution of using random forest classifier algorithm to detect and predict detection. A mobile application will be developed in order to collect user data and make prediction.

KEYWORDS

Depression; machine learning; algorithm

INTRODUCTION

“According to (Manar & Stephane, 2015) machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence”. Machine learning algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions, rather than following strictly static program instructions. Machine learning is closely related to and often overlaps with computational statistics; a discipline that also specializes in prediction-making. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. One of the biggest problems in treating patients is that drugs often affect individuals differently. Some medications may

cause terrible side effects for one patient while being an effective treatment for a different patient. A patient may have additional medical conditions that may cause a reaction to a treatment. One solution to selecting the most effective treatment is to build a machine learning model based on classification and regression algorithms. The classification model is needed to predict the impact of the drug based on known results from patient tests and conditions. The regression model is then used to predict the changes in the patient's condition when she takes a certain drug. Creating this model by using data helps provide researchers with an understanding of how a population of patients historically reacts to various drugs. As the model is built and trained, it will be able to determine the probability that a certain drug will be most effective for a patient. If the model is online, it will continue to evolve as more patient data is added. A solution can be built to include a conversational interface using cognitive Application Programming Interfaces (APIs). In this way, a physician can interact with the model and ask a variety of questions to ensure that the right treatment is provided with fewer side effects (Hurwitz & Kirsch, 2018).

RESEARCH QUESTIONS

What main environment, behavioral and family factors predict whether Africans will develop depression and how can these commonalities be used to aid the medical community in diagnosing early depression?

What are the machine learning algorithms available and why will this be more appropriate in this context?

What significant impact of mobile application technology in diagnosing depression using machine learning?

DATA COLLECTION

In this work, data from a concluded Artificial Intelligence competition in Zindi Africa. The Competition was a one day hackathon hosted by the Busara Center for Behavioral Economics in Nairobi, it was based on routine survey. These data came from a 2015 study conducted by the Busara Center in rural Siaya County near Lake Victoria in western Kenya. The survey includes more than 70 features, including information about household composition, economics activity, financial flows and health.

DATA REDUCTION

Only 15 features will be chosen from the features in the data set. These features includes:

- Survey date
- Gender
- Age
- Marriage status
- Children
- Imitate household size
- Number of years of education
- Total household size
- Expenses on food (last month)

- Medical expenses (last month)
- Expenses on education (last month)
- Expenses on social activities (last month)
- Depressed. The data set will be split into train and test data set.

DATA CLEANSING

Random forest Classifier algorithm does not support null values. Therefore to execute random forest algorithm, null values have to be managed from the original raw data set. Dropping or deleting rows with null values will be implemented on the data set.

RESULT INTERPRETATION

A probability closer to 1 signifies depressed while a probability closer to 0 signifies not depressed.

DEPLOYMENT

The model will be deployed to a cloud server and a mobile application will be developed in order to collect user data and make prediction. The mobile app and the deployed model will communicate via an API (application programming interface).

CONCLUSIONS

In conclusion, the implementation is yet to be carried out, with the hope that the detection and prediction of depression mobile application will be made available to health workers. This will help them in monitoring patients.

REFERENCES

- [1] Manar, A. and Stephane, P. (2015). Machine Learning with Python/scikit-learn: applications to the Estimation of occupancy and human activities. SUMMUREX 2015. pp.5-8
- [2] Judith Hurwitz and Daniel Kirsch (2018). Machine learning for dummies. IBM limited edition. Published by John Wiley and Sons Inc. Hoboken, New Jersey, USA
- [3] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in Proceedings of ICWSM, 2013, pp. 1–10.

- Expenses on alcohol (last month)
- Expenses on tobacco (last month)

failure,” Primary Care Companion to the Journal of Clinical Psychiatry, vol. 11, no. 3, pp. 103–109, 2009.

[5] L. E. Egede, “Failure to recognize depression in primary care: issues and challenges,” 2007.

[6] Mental health and older adults. Available at: <http://www.who.int/mediacentre/factsheets/fs381/en/>, accessed 12 November 2016

[7] Mathur R., Pérez-Pinar M., Foguet-Boreu Q., ET AL.: ‘Risk of incident cardiovascular events amongst individuals with anxiety and depression: a prospective cohort study in the East London primary care database’, J. Affect. Disord., 2016, 206, pp. 41–47

[8] Watkins L.L., Koch G.G., Sherwood A., ET AL.: ‘Association of anxiety and depression with all-cause mortality in individuals with coronary heart disease’, J. Am. Heart Assoc., 2013, 2, (2), p. e000068

[9] ‘WHO|Depression’, WHO. Available at: <http://www.who.int/mediacentre/factsheets/fs369/en/>, accessed 12 November 2016

[10] Djernes J.K.: ‘Prevalence and predictors of depression in populations of elderly: a review’, Acta Psychiatr. Scand., 2006, 113, (5), pp. 372–387

[4] J. A. Cully, D. E. Jimenez, T. A. Ledoux, and A. Deswal, “Recognition and treatment of depression and anxiety symptoms in heart

