CLINICAL DECISION SUPPORT TOOL FOR DETECTING MENTAL HEALTH CONDITION OF CHILDREN

Adhri Nandini Paul Lakshman B. Boddu Vishal Sheshadri

Department of Computer Science
Lakehead University

***Abstract*—Early diagnosis of mental health problems will help professionals to treat it at an earlier stage and improves the patients' quality of life. So, there is an urgent need to treat basic mental health problems that prevail among children which may lead to complicated problems, if not treated at an early stage.**

This project is about predicting mental health disorder in children. The project will ask the user(in our case, a parent) a set of questions on how the mental health of the child is. Depending upon the answers, a prediction to whether the child has mental health problem or not is generated. We have calculated the performance measures of the classification models, i.e, Specificity and Sensitivity. A data set consisting of 200 records is collected for training and testing the performance of the techniques. Ten attributes have been identified as important for diagnosing the problem from the documents. The attributes have been reduced by applying Feature Selection algorithms over the full attribute data set. The accuracy over the full attribute set and selected attribute set on various machine learning techniques have been compared.

Keywords— Child Mental Health Diagnosis; Feature Selection; Child Mental Health Problems

I. INTRODUCTION

Childhood and the early years of adolescence are a time of life when many changes occur, for example starting school, leaving home and meeting new kids. For many, these are exciting times. They can also be times of stress and apprehension however. In some cases, if not recognized and managed, these feelings can lead to mental illness. Many people are also living in areas affected by humanitarian emergencies such as conflicts, natural disasters and epidemics. Young people living in situations such as these are particularly vulnerable to mental distress and illness. On mental health day this year(2018), World Health Organization (WHO) stress the importance of "Young People and Mental Health in a changing world"[1]. According to WHO, "half of all

mental illness begins by the age of 14, but most cases go undetected and untreated".

Detecting these mental illnesses early can help in improving the mental health of children. They can grow up to be the person the society accepts, rather than being kept away from the society.

Mental health diagnosis involves many steps and it is not a straightforward process. Actually, the diagnosis starts with specially designed interview with questions about the symptoms and medical history and sometimes performing a physical examination. Various psychological tests are also conducted to make sure that the symptoms are caused only by the mental health and not by any other problems. Various assessment tools are available to evaluate a person for a mental illness.

Our project is a similar assessment tool that asks a set of questions and then uses machine learning algorithms[2] to determine whether the child has a problem or not, depending on the previous cases used in training the system.

II. OVERVIEW OF MACHINE LEARNING AND ASSESSMENT TECHNIQUES

There is no single learning machine algorithm that universally performs best across all domains. Hence, in this project the following techniques are selected as they produced near-accurate results for a small dataset. There are brief description about these algorithm.

A. Logistic regression

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is having binary values. Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

B. K-Nearest Neighbours Algorithm (k-NN)

K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The *k*-NN algorithm is among the simplest of all machine learning algorithms. Both for classification and regression, a useful technique can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones.

C. Principal Component Analysis (PCA)

PCA simplifies the complexity in high-dimensional data while retaining trends and patterns. It does this by transforming the data into fewer dimensions (attributes), which act as summaries of features. High-dimensional data are very common and arise when multiple features are measured for each sample. This type of data presents several challenges that PCA mitigates: computational expense and an increased error rate due to multiple test

correction when testing each feature for association with an outcome. PCA is an unsupervised learning method and is similar to clustering — it finds patterns without reference to prior knowledge about whether the samples come from different treatment groups or have phenotypic differences.

D. K-Fold Cross Validation

In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k − 1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once.

III. METHODOLOGY

As the first step for the project, we diagnose basic mental health by taking providing a questionnaire[2][3] to identify the mental health problems that occur more often among children. Then, observation is made on how the answers were provided to the questions. A model is built that uses machine learning techniques to predict the mental health of the children. This model assists the professionals to identify the problem if the known evidences of the patient are given as input.

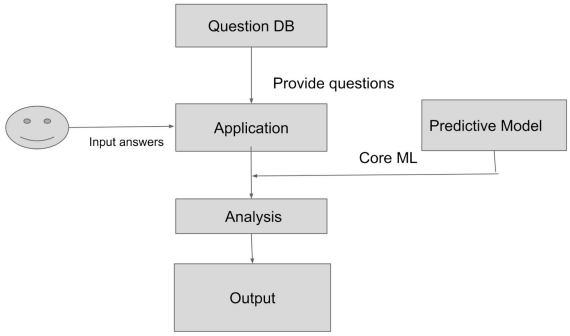


Fig. 1. Methodology of the Project

The questions are stored in a model. The questionnaire application extracts those questions and populates as a questionnaire to the user. The user then answers the questions as per the situation of the child.
The CoreML[4] machine learning libraries in XCode software to link the machine learning model to the application. The predictive model will take the existing dataset and analyzes it to predict if the child will have any kind of disorder depending upon the answers. To achieve desired results, dataset consisting of empirical data is to be used.[5]

IV. MODEL EVALUATION

To train child mental health data, we has chosen Logistic Regression and K-nearest Neighbourhood algorithm. Logistic regression is used to calculate the accuracy of outcome of disease when it is provided with different attributes, i.e, those discussed in the dataset.
k-NN classifier is also used to do the same. It calculates and produces the accuracy of our system from the attributes. Afterward, we have used k-fold cross validation in

order to estimate the efficiency of making prediction during training and testing the dataset. We calculate the Sensitivity and Specificity, and plot the Receiver Operating Characteristic (ROC) Curve for the data.

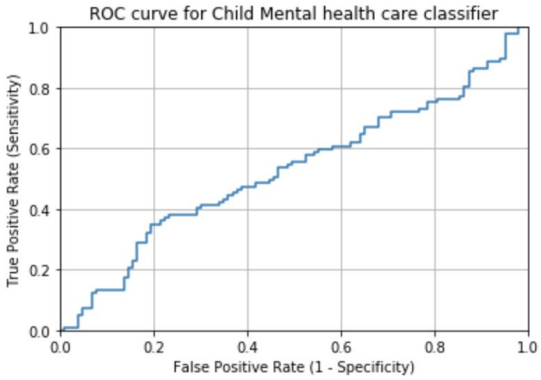
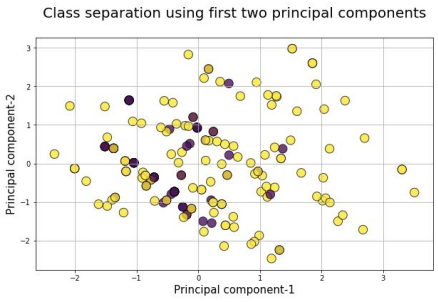


Fig. 2. ROC Curve for Sensitivity vs (1-Specificity)

We have removed a few attributes during calculation, since they are not assisting in the prediction of disease.

To reduce dimension, we have used Principal Component Analysis(PCA). After fitting the data with PCA, the variance of first two components were close. The 1st principal component explains about 16% of the total variance in the data and the 2nd component depicts next 14%. Therefore, if we just consider first two components, they together explain 30% of the total variance. By plotting the data for these two components, patients and non-patients are clustered independently.



Our model is giving an accuracy of 68.55% when evaluated using logistic regression and an accuracy of 72.80% when k-nn classifier is used.

V. CONCLUSION

Nowadays, a number of decision support systems are used in the medical field. But still there is a lot of scope of improvement and upgradation required in this field.

This project might help the psychiatrist or the mental health practitioner to predict the mental health condition of a child. But it needs improvement in the areas of accuracy and dependability. This can be achieved by using empirical data (real dataset).

VI. FUTURE WORK

There are some more features that could be added to this project to improve the efficiency of the project and provide better accuracy. As, we have worked with labeled data by following supervised learning algorithm, in future we would use autoencoder instead of PCA for dimension reduction which supports unsupervised learning algorithm to utilize more features of data. Moreover, the following methods can be used to enhance user experience.

- Add more disorders and functionality to the application.
- Work on real-life dataset.
- To develop a messaging system that informs the doctor if a patient is suffering from severe disorders and take continuous guideline.

VIII. SOURCE CODE

The project source code can be found at: https://github.com/lakshmanboddu/ChildMentalProblemPrediction.

The dataset and evaluation can be found at: https://github.com/Adhrimuna/MentalProblemPrediction

IX. ACKNOWLEDGEMENTS

This project report is submitted as a requirement partial fulfillment for the completion of the course COMP-5413 at Lakehead University during Fall Term of the year 2018.

We thank our professor Dr. Joshua J. Armstrong, PhD., from Lakehead University who provided insight and expertise that greatly assisted the project, although he may not agree with all of the conclusions of this project.

We thank Ms. Sirisha Peram, for assistance with identifying defects and providing valuable inputs that greatly helped in improving the project.

We would also like to show our gratitude to the Lakehead University for providing their infrastructure and support to us during the course of this project. We are also immensely grateful to our classmates for their comments during earlier versions of the project. Any errors present are our own and should not tarnish the reputations of these esteemed persons or entities.

VII. REFERENCES

[1] WHO Mental Health Day Article, 10.10.2018 http://www.who.int/mental_health/world-mental-health-day/2018/en/

[2]Sumathi M.R., Poorna B. (2016) "Prediction of Mental Health Problems Among Children Using Machine Learning Techniques", *International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, 2016.* http://thesai.org/Downloads/Volume7No1/Paper_76-Prediction_of_Mental_Health_Problems_Among_Children.pdf?fbclid=IwAR2t_5EMNx2Qrzrlr6Bz5bW3I2kkOAA-6JXoUHS5TzbgO30cCg6Nb5ms3oI

[3] Doenlen, MD, Henry A. (1986) "Decision Trees for Use in Childhood Mental Disorders," Jefferson Journal of Psychiatry: Vol. 4 : Iss. 1 , Article 9. DOI: https://doi.org/10.29046/JJP.004.1.006 https://jdc.jefferson.edu/jeffjpsychiatry/vol4/iss1/9

[4]CoreML: https://developer.apple.com/documentation/coreml

[5] Haiwens S., James L.W. (2007) "Clinical Decision Modelling System" *BMC Medical Informatics and Decision Making* 2007 **7**:23 https://doi.org/10.1186/1472-6947-7-23

--x--