Towards the Analysis and Prediction of Early Childhood Development Using Data Mining Classification Techniques

[1]I. Aimufua, [2]S. Rakshit, [3]N.R. Vajjhala & [4]O.B. Longe
School of IT & Computing
American University of Nigeria, Yola, Nigeria
ikponmwosa.aimufua@aun.edu.ng; sandip.rakshit@aun.edu.ng narasimha.vajjhala@aun.edu.ng;
olumide.longe@aun,.edu.ng

Abstract - In recent years, the use of technology in almost every sector has been on the increase, which has lead to massive growth in the generation and usage of data. Medical, Business and other leading industries have surveyed and noted that data repositories will be a useful tool in the designing of business strategies, analyzing of unstructured and structure data in other to gain useful knowledge.

Keywords: Early Childhood Development Analysis and Prediction Using Data Mining Classification Techniques

I. INTRODUCTION

In the educational domain, especially the early childhood education, there are very rich information available that we could leverage on to determine the development stage of a child and hidden patterns of a child learning ability or disability, but we are faced with weak knowledge and tools which are needed to harness this potentials. This study focuses on investigating which data mining classification technique the will be most suitable in building a predictive model that can identify the social, cognitive and emotional stage of a child. Through massive literature review, it was found that EDM has started gaining traction as researcher are now integrating learning management system(LMS) with teaching module in order to gain useful insight. As the education data mining get larger into early childhood education by so doing having different characteristics, different data mining technique will have their own predictive efficiency [1].

In this study, we will comprehensively study and compare different data classification technique and their prediction accuracy for early childhood dataset. Authors compared J48, Naïve Bayes, Random Forest, SVM and k-NN classifiers using performance measure like kappa statistics, ROC, RMSE and MAE using a data mining analytical tool called WEKA. Author also compared their accuracy measure like TP rate, FP rate, precision, recall and F measure. The result show from all analysis indicate that J48 classifier has a better classification accuracy and prediction rating over other tested algorithm using early childhood data set.

A. Background of Study

Early childhood can be classified as the stage where major transition in infant occur which is a very significant influence on the development of the child and the transition to middle childhood stage, late childhood stage and then adolescence. According to [3] early childhood education should aim at holistic development of a child's Social, emotional, cognitive and physical need in other to build a foundation for lifelong learning and wellbeing.

In low and middle income countries, there are more than 200 million children under the age of 5 that have not gotten to their full developmental potential due to certain factors like poverty, poor nutrition and other factors [2].

Research has proven that the experiences of a child in their early child hood have a long term effect on the psychological, cognitive and general brain function.

Also, the exposure of a child to an environment that is not so conducive for upbringing during the first few years of their life lowers the child intelligent quotient (IQ), can also lead to poor academic achievement, increased anti-social behaviours and lowers economic productivity in their adulthood [4].

A child's characteristic of behaviour and self control in his developmental stage (age 1-3) predicts psychopathology [5]. Hence, the proper development of early childhood is the physical, mental, behavioural, cognitive and spiritual foundation of a well productive adulthood.

In [6] factors which are associated with resilience in early childhood development were listed listed to be the connection to competent and caring adult, cognitive and self regulation skill, a self positive image and finally motivation. The presence of these factors in early child development can influence several aspect of an overall child's functioning, which include the social behaviour, the academic performance and metal health

Data Mining is a multi disciplinary field that is gaining a lot of focus as it is used to know or identify hidden information or pattern from a large dataset. The field of data mining has gotten a whole lot of traction because of its analytical capabilities as it is been applied to several research domain which includes Medical prediction and Health care analysis, Social Network Mining, Business analysis/Intelligence and Education [7]

In the educational domain (early childhood education) there are several rich information available but we are faced with a very weak knowledge and tools which are needed to identify hidden pattern for early childhood development [9] [9]. Data Mining technique can be applied to the field of Education especially the early childhood education to help build predictive models that will be able to identify the current developmental stage such as the social, cognitive and emotional stage of a child and the support the child needs and be able to build systems for early prediction and correction of behavioural pattern of a child

Data Mining is a field in computer science that involves several computational processes, artificial intelligence, statistical machineries and tools which is used for extracting useful information and for pattern discovering. Mining data from early childhood education can lead to better understanding on the teaching method a child need, Data mining could be used in developing a predictive model for the educational ecosystem that collects, process and review hidden information or pattern of a child which can help in predictions and classification's of childhood challenges to better improve the early childhood developmental process.

II. PROBLEM STATEMENT

A. Problem Statement

Most study on data mining has always focused on higher level of education, medical analysis and prediction and lastly business prediction model development and analysis, but little or no study have been done in developing and analyzing a predictive model for early childhood development dataset.

This study will focus on evaluating different classifiers technique with different key performance indicators targeted at early childhood development dataset to determine which classifier has a better prediction and a more accurate performance in developing a data mining model for early childhood development [10].

III. RESEARCH QUESTION

In trying to understudy the educational data mining of early childhood, we came up with this approach on

- Which classification technique (J48, Random Forest, Naïve Bayes, SVM, and KNN) will be most appropriate in developing a predictive model for early childhood education system in Nigeria?
- What are the relevant factors that influence early childhood development in Nigeria?

IV. RESEARCH THRUST

The research aims to analyze and predict the early childhood development using data mining classification technique. The objective is to investigate which data mining algorithm is more accurate in developing a data mining model for early childhood education. The specific objective of this proposal includes

- To investigate how close the classification technique predictions is to possible actualization
- To investigate the technique which are suitable to predict classes of needed attention in Early childhood education dataset
- Implementation of the predicted model algorithm using WEKA to analyze Early childhood education dataset

A. Scope of Study

The emphasis of this research will be to identify the classification technique and analyze the algorithm best suited for model development given a specific dataset of early childhood education. We will be using some indicators as performance measurement like Accuracy for value prediction, MAE (Mean Absolute Error) to check how close the prediction is to possible outcome, RMSE (Root Mean Squared Error/Deviation) to differentiate the predicted value and observed value, K-statistics to compare the accuracy of the system to random numbers and ROC (Receiver Operating Characteristics) for graphic representation of performance.

B Justification of Research

The Justification of this research is to exploit classification algorithm available in data mining to provide an accurate and precise performance rating for each algorithm given a dataset for early childhood development. Investigating algorithm with best performance will provide insightful knowledge for data mining model development for early childhood education in Nigeria

V. RESEARCH AGENDA

We will review literature on data mining analysis on two different research domain; medical domain and educational domain. Application of data mining classification technique for the virtualization of hidden pattern and the comparison of different classifiers to indicate which have a better accuracy will be discussed in this chapter. Furthermore, the research gap, research framework underpinning data mining in the educational domain will also be examined.

A detail of the research method to be used will be outlined. Also, the research approach, the theoretical background of the thesis will be extensively discussed. The method in which data is to be collected, the attribute needed and the tool that is needed to analyze the data will be discussed. We will then proceed to the descriptive analysis of the data collected. The experimental result from the analysis is to be extensively discussed. We will then come out with the result obtained. Also, the chapter concludes and gives recommendation for further research.

REFERENCES

{1}   Agresti, A., & Finlay, B. (2009). Statistical Methods for the Social Sciences. Pearson Prentice Hall.

[2]   Alaa, T. (n.d.). Classification assessment methods | Elsevier Enhanced Reader. https://doi.org/10.1016/j.aci.2018.08.003

[3]   Ambili, & Afsar. (2016). A Framework for Child Development analysis and Learning Disability Prediction using a Hybrid Naive Bayes and Decision Tree Fusion Technique – NB Tree. https://www.semanticscholar.org/paper/A-Framework-for-Child-Development-analysis-and-a-%E2%80%93-Ambili-Afsar/13b46ecaf0fb49eec19cb31ddbe2228053739e3b

[4]   [5]   Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, 60(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

[6]   Ankita, N., & Anjali, R. (2017). Analysis of Student Performance Using Data Mining Technique. www.ijircce.com

[7]   Aqra, I., Herawan, T., Ghani, N. A., Akhunzada, A., Ali, A., Razali, R. B., Ilahi, M., & Choo, K.-K. R. (2018). A novel association rule mining approach using TID intermediate itemset. PLOS ONE, 13(1), e0179703. https://doi.org/10.1371/journal.pone.0179703

[8]   Baker-Henningham, H. (2014). The role of early childhood education programmes in the promotion of child and adolescent mental health in low- and middle-income countries. International Journal of Epidemiology, 43(2), 407–433. https://doi.org/10.1093/ije/dyt226

[8]   Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. Archives of General Psychiatry, 53(11), 1033–1039. https://doi.org/10.1001/archpsyc.1996.01830110071009

[9]   Chakrabarty, A., Mannan, S., & Cagin, T. (2016). Chapter 8—Inherently Safer Design. In A. Chakrabarty, S. Mannan, & T. Cagin (Eds.), Multiscale Modeling for Process Safety Applications (pp. 339–396). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-12-396975-0.00008-5

[10]   Chen, M.-S., Han, J., & Yu, P. S. (1996). Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866–883. https://doi.org/10.1109/69.553155