

A Framework for Child Development analysis and Learning Disability Prediction using a Hybrid Naive Bayes and Decision Tree Fusion Technique – NB Tree

Ambili K ¹, Afsar P²

P.G. Student, Department of Computer Engineering, M.E.A Engineering College, Perinthalmanna, Kerala, India¹

Assistant Professor, Department of Computer Engineering, M.E.A Engineering College, Perinthalmanna, Kerala, India²

ABSTRACT: Child development analysis has long been a research interest that seeks to understand and explain the different aspects of growth, including physical, emotional, intellectual, social, perceptual and personality development. By better understanding how and why people change and grow, one can apply this knowledge to understand the needs of a child and fulfilling them and allow them to reach their full potential. Clearly, the aim of child development is broad and the scope of the field is extensive. The research study therefore focus to apply a datamining approach to predict the child's learning behavior and skills using machine learning algorithms. The information gained from this evaluation is further used to predict learning disability found in children. The purpose of learning disability prediction is to determine a child's strengths and weaknesses so that parents and school authorities can provide the best learning environment for a child. The observations show that the prediction model developed using a Hybrid Naive Bayes and Decision Tree fusion technique-NB Tree is found to be the best among classification and prediction algorithms for child development analysis and learning disability prediction.

KEYWORDS: Child development analysis, Learning disability, Naive Bayes, Decision Tree, NB Tree.

I. INTRODUCTION

Child development analysis has long been a research interest that seeks to understand and explain the different aspects of growth, including physical, emotional, intellectual, social, perceptual and personality development. In order to study the growth, change and stability, child development takes a scientific approach. Researchers in child development test their assumptions about the nature and course of development by applying scientific methods. The study of child development is important in a number of fields including Biology, Anthropology, Sociology, Education, Psychology, Pediatrics etc.. However most important are the practical applications of studying child development. By better understanding how and why people change and grow, one can apply this knowledge to understand the needs of a child and fulfilling them and allow them to reach their full potential. Evidence tells that a person's life successes, health and emotional wellbeing have their roots in early childhood. The quality of a child's earliest environments and the availability of appropriate experiences at the right stages of development are crucial determinants of the way each child's brain architecture develops.

A learning disability can cause a child to have some kind of trouble in learning and using certain skills. The skills most commonly affected are reading, writing, listening, speaking, reasoning and doing math. Learning disability prediction is a very complicated task. The purpose of prediction of learning disability during early childhood is very essential since it helps to determine child's strengths and weaknesses and to understand how he or she best learns and where they have difficulty. The information gained from an evaluation is very important for finding out how the parents and the school authorities can provide the best learning environment for child.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

A major idea of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. In recent years the sizes of databases have increased rapidly. This has lead to a growing interest in the development of tools capable in the automatic extraction of knowledge from data. The term Data Mining or Knowledge Discovery in databases has been adopted for a field of research dealing with the automatic discovery of implicit information or knowledge within databases. Knowledge Discovery in Databases (KDD) is the process of identifying useful information in data . Data mining is the non-trivial extraction of implicit previously unknown and potentially useful information about data. Conventionally, the information mined is denoted as a model of the semantic structure of the datasets.

Clearly, the aim of child development is broad and the scope of the field is extensive. However, only a limited number of studies have been focussed in this field. The application of datamining in early childhood research is still at an infancy stage and is yet to be widely embraced. The prediction model for child development analysis covers quite diverse areas.

This paper is organized as follows. Section II presents some related work and recent studies on child development analysis using data mining techniques. Section III gives a brief overview of the available data and the transformations carried out to clean and put the data in the proper format for analysis. Section IV gives the description of the proposed approach which have shown best accuracy with our dataset. Section V presents the obtained results and Section VI concludes with some remarks about the described work.

II. RELATED WORK

The application of data mining in early childhood research is still at an infancy stage. There are only very limited studies conducted on the adoption of data mining techniques in analysing early childhood datasets[1]. Clearly, the aim of child development is broad and the scope of the field is extensive.

Finally, child development focuses on the ways people change and grow during their lives. It seeks in which areas and in what periods, people show change and growth and when and how their behaviour reveals consistency and continuity with prior behaviour. Some of the data mining technique used for child development analysis used machine learning algorithms such as Rough set approach and Decision tree algorithm, Fuzzy expert systems, Neural Networks etc..[2,3,4]. The Rough set approach seems to be of fundamental importance to artificial intelligence [5,6]. Rough set theory (RST) has been successfully applied in many real life problems such as medicine, pharmacology, engineering, banking, finance, market analysis, environment management and others. The rough set approach of data analysis has much important advantage.

During the late 1970s and early 1980s, J. Ross Quinlan, a researcher in machine learning developed a decision tree algorithms known as ID3 [7]. This work expanded on earlier work on concept learning system. This work expanded on earlier work on concept learning system. Decision tree method is widely used in data mining and decision support system. Decision tree is fast and easy to use for rule generation and classification problems. It is an excellent tool for decision representations. The accuracy of a classifier refers to the ability of a given classifier to correctly predict the class label of new or previously unseen data.

For prediction of learning disability, decision trees are probably the most frequently used tools for rule extraction from data,[7,8] whereas the rough sets based methods seems to be their newer alternative. In both cases, the algorithms are simple and easy to interpret by users. The practical aspects of application of those tools are different. The computational times of decision trees are generally short and the interpretation of rules obtained from decision trees can be facilitated by the graphical representation of the trees.

RST may require long computational time and may lead to much large number of rules compared to DT[9]. The rules extraction algorithm is very important, particularly in construction of data mining system. Therefore, we have to go for some other machine learning algorithms.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

III. DATA COLLECTION

The data set used for the research focus on information regarding various milestones of child development in all perspectives. It covers quite diverse areas including physical development, cognitive development, knowledge and understanding of the world, communication, language and literacy, personal-social and emotional development, problem solving, reasoning, numeracy and creative skills. The primary methods for collecting data are interviews and questionnaires. The child development data was collected from various sources including psychologists, school councellors, MSW child welfare workers, parents, websites, and books related to Child development and pedagogy, Advanced pediatric assessment etc.

Data Analysis : - Participants in the research are parents/ caretakers/ teachers of children aged between 0-8 years. The purpose of the research and brief data collection process are explained to them.

Data Preparation:- Age and domain related questionnaire is prepared based on the different domains of child development. The questionnaire contains statements concerning the skills and behaviours of children in various domains of development. The statements in the questionnaire are followed by boxes marked "Does not apply", "Applies sometimes" or "Applies". The parents have to respond to the questionnaire by choosing a box that contains the statement that they think best corresponds to their child's functioning in everyday situations.

Data Selection and Transformation:- The useful information is selected according to requirements and the data in pdf format will be converted to rtf format using miscellaneous tools and tricks. Data preprocessing is done to handle missing values, noise and outliers.

Input Variables:- From the vast initial dataset, a limited number of important attributes are selected which have the highest contribution to analyse the developmental factors. These attributes are however age dependent. They are :-

1. Gross Motor
2. Fine Motor
3. Communication
4. Problem Solving
5. Personal, Social And Emotional Development

Table 3.1:- Factors analysed for age 0 to 5 years.

1. Gross Motor	2. Perception Of Space
3. Fine Motor	4. Concept of Time
5. Communication	6. Memory
7. Problem Solving	8. Forms And Figures
9. Personal, Social	10. Spoken Language
11. Attention, Concentration	12. Academics In School
13. Overactivity, Impulsivity	14. Reading, Writing, Math
15. Passivity, Inactivity	16. Social Skills
17. Planning, Organising	18. Emotional Problems

Table 3.2:- Factors analysed for age 6 to 9 years.

Now, to diagnose learning disability, a checklist of 16 factors are used. It contains 16 most frequent signs and symptoms of LD. These symptoms, which are the attributes in this study, are listed in table below.-

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

SL NO	ATTRIBUTES	SYMPTOMS OF LD
1	DR	Difficulty with Reading
2	DS	Difficulty with Spelling
3	DH	Difficulty with Handwriting
4	DWE	Difficulty with Written Expression
5	DBA	Difficulty with Basic Arithmetic
6	DHA	Difficulty with Higher Arithmetic
7	DA	Difficulty with Attention
8	ED	Easily Distracted
9	DM	Difficulty with Memory
10	LM	Lack of Motivation
11	DSS	Difficulty with Study Skills
12	DNS	Does Not like School
13	DLL	Difficulty in Learning a Language
14	DLS	Difficulty in Learning a Subject
15	STL	Slow To Learn
16	RG	Repeated a Grade

Table 3.3:- Factors analysed in diagnosis of LD

This data was used as the training set for various algorithms. The testing data was collected through the questionnaire of 30 school children.

IV. PROPOSED METHOD – NB TREE MODEL

The framework for predicting child development analysis uses a Hybrid Naive Bayes and Decision Tree technique. Both these algorithms are good classification and prediction techniques individually. By combining both these techniques, more accurate prediction techniques can be obtained. The architecture of the prediction model is shown in figure 1.

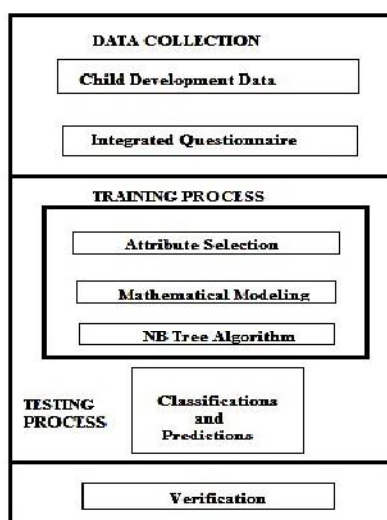


Figure 4.1: The prediction framework architecture

There are five modules in the proposed framework. They are :

- Data Collections

International Journal of Innovative Research in Science, Engineering and Technology

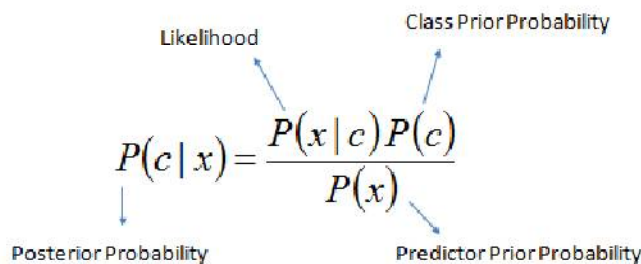
(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

- Child Development factors identification and Modelling
- Classifications and Predictions
- Ld prediction
- Verification

Classifications And Predictions:- The NB-Tree technique is a hybrid of two classifiers :- the ID3 Decision Tree and Naive Bayes. ID3 is interesting in its representation of knowledge, its approach to the management of complexity, its heuristic for selecting candidate concepts, and its potential for handling noisy data. It represents the concept of decision tree, that allow for classification for an object by testing its value for certain properties. The Naive Bayes classifier is based on the Bayesian theorem and is particularly suited for high dimension inputs. It is simpler than most methods but it still outperforms other sophisticated classification techniques.

Naive Bayes Formula:- The naive Bayes classier greatly simplify learning by assuming that features are independent of given class. Naive Bayes model records how often a target field value appears together with a value of an input field. It considers each of the symptoms to contribute independently to the probability that the child has proper development or not. It estimates the probability of observing a certain value in a given class



$$P(c | x) = \frac{P(x | c) P(c)}{P(x)}$$

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Figure 4.2:- Naive Bayes Formula

Algorithm Decision Tree:-The Decision Tree is a ow chart like structure, where each internal node denotes a test on an attribute, each branch of the tree represents an outcome of the test and each leaf node holds a class label. The topmost node in a tree is the root node. Decision trees are powerful and popular tool for classification and prediction. It is a classifier in the form of a tree structure where each node is either a leaf node that indicates the value of the target attribute of examples or a decision node that specifies some test to be carried out on a single attribute with one branch and sub tree for each possible outcome of the test. Classifiers do not require any domain knowledge or parameter setting and therefore is appropriate for exploratory knowledge discovery. Decision tree can handle high dimensional data. The learning and classification step of decision tree are simple and fast. A decision tree can be used to classify an example by starting at the root of the tree and moving through it until a leaf node, which provides the classification of the instance.

A tree building process starts by selecting an attribute to place at the root node and at each succeeding level the subset generated by proceeding levels are further partitioned until it reaches a relatively homogeneous terminal node or leaf node.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

The algorithm for ID3 Decision tree is shown below:

```
function induce_tree (children_set, DevptFactors)
begin
if all entries in children_set are in the same class
then return a leaf node labeled with that class
else if Devpt_Factors is empty
then return leaf node labeled with disjunction of all classes in children_set
else begin
select a property, P, and make it the root of the current tree;
delete P from DevptFactors;
for each value, V, of P,
begin
create a branch of the tree labeled with V;
let partition V be elements of children_set with values V for property P;
call induce_tree (partition, DevptFactors), attach result to branch V
end
end
end
```

Figure 4.3:- Algorithm ID3 Decision Tree

The condition attribute, that induces most amount of entropy reduction and information gain are placed closer to the root node.

V. EXPERIMENTAL RESULTS

The experiment makes a comparative study on the performances of machine learning algorithms for child development analysis. They are evaluated on the basis of three criteria :-

1. Prediction Accuracy
2. Learning Time and
3. Error Rate

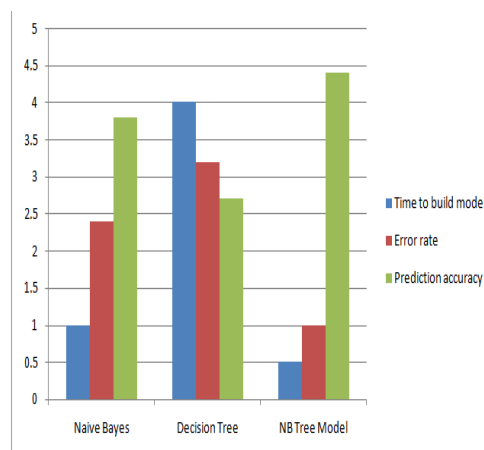


Figure 5.1:- Performance comparison of different models

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

From the results, it was able to understand that our proposed approach NB Tree algorithm provides more number of correctly classified instances than the other two algorithms. Regarding the Learning time of algorithms, it was able to understand that Decision Tree model consumes more time to build the model. Out of these three algorithms, our proposed method has high prediction accuracy than other two algorithms.

VI. CONCLUSION

In this research study, a comparative study was conducted on various datamining classification and prediction algorithms for child development analysis. The framework for predicting child development analysis uses a Hybrid Naive Bayes and Decision Tree technique. Both these algorithms are good classification and prediction techniques individually. By combining both these techniques, more accurate prediction techniques can be obtained. From the study it was able to conclude that the proposed framework out performs other machine learning algorithms in terms of prediction accuracy, time consumption and error rate.

REFERENCES

- [1] M. Gera and S. Goel, "A model for predicting the eligibility for placement of students using data mining technique," International Conference on Computing, Communication and Automation, vol. 4, pp. 18-23, January 2015.
- [2] Sellappan Palaniappan and Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," IJCSNS International Journal of Computer Science and Network Security, Vol.8, No. 8, August 2008.
- [3] E. A.Q. Ansari, Neeraj Kumar Gupta, "Automatic diagnosis of asthma using neurofuzzy system," Fourth International Conference on Computational Intelligence and Communication Networks, Vol.7, April 2012.
- [4] Muhamad Hariz and Wahidah Husain "A Framework for Childhood Obesity Classifications and Predictions using NBtree," International Conference on IT in Asia, No.8, November 2011.
- [5] Cios, K.J., Pedrycz W., Swiniarski, R.W. and Kurgan, L.A., "Data Mining: A Knowledge Discovery Approach", Springer, New York.
- [6] Grzymala-Busse JW, " Knowledge Acquisition under Uncertainty-A Rough Set Approach," Journal of Intelligent & Robotic Systems, Vol 1, 3-16, 1988.
- [7] Quinlan J.R., "Induction on decision trees, Machine learning," Vol 1, 81-106, 1986.
- [8] Stuart R., Peter N., "Artificial Intelligence – A Modern approach," Pearson Prentice Hall, 2009.
- [9] Julie M. David, Kannan Balakrishnan, "Prediction of Frequent Signs of Learning Disabilities in School Age Children using Association Rules," In Proceedings of the International Conference on Advanced Computing, Vol 13, April 2009.
- [10] Chen R.S, Wu R.C., Chang C.C, "Using data mining technology to design an intelligent CIM system for IC manufacturing," in proceedings of sixth international conference on software engineering, artificial intelligence, network, parallel distribution, computation self assembly wireless network, 70-75, 2005.
- [11] Kusiak A, Kurasek C, "Data mining of printed circuit board defects", IEEE Transactions on Robotic Automations, Vol 17, No.2, 74-84, March 2006.
- [12] Tseng T.L., Jothishankar M.C., Wu T., Xing G., Jiang F, "Applying data mining approaches for defect diagnosis in manufacturing", World Academy of Science, Engineering and Technology, 61, 2010.