International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 07 | July-2016 www.irjet.net p-ISSN: 2395-0072

A Prediction Model for Child Development Analysis using Naive Bayes and Decision Tree Fusion Technique – NB Tree

Ambili K[1], Afsar P[2]

[1]*M.Tech Student, Dept. of Computer Science & Engineering, MEA Engineering College, Perinthalmanna*

[2] *Assistant professor, Dept. of Computer Science & Engineering, MEA Engineering College, Perinthalmanna*

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. Inorder to study the growth,change and stability, child development analysis takes a scientifc approach. By better understanding how and why people change and grow, one can apply this knowlede to understand the needs of a child and fulflling 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. However, only a limited number of studies have been focussed on the field of early childhood development. The research study therefore focus to apply a datamining approch to predict the child's future learning behavior and skills using machine learning algorithms. Here, the prediction model is developed using a Hybrid Naive Bayes and Decision Tree fusion Technique - NB Tree.*

Key Words: Child development, Naive Bayes Algorithm, Decision Tree, Hybrid Naive Bayes – Decision Tree Algorithm.

1.INTRODUCTION

Child development is the field that involves the scientifc study of the patterns of growth, change and stability that occur from conception through adolescence. It gives an understanding of how a child is able to do complex things as he gets older. The study of child development is important in a number of felds 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 knowlede 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 well being have their roots in early childhood. The quality of a child's earliest environments and the availability of appropriate experiences at the right satges of development are crucial determinants of the way each child's brain architecture develops.

1.1 Aims of Child Development

- To make us aware that the child is developing normally.
- To enable us to identify a child, who for the some reason, may not be following the normative stages.
- To enable us to build up a picture of a child's progress for a particular period of time.
- To help us to consider the fact that every child is different from each other in quite normal ways.
- To make us aware that every child follows the same sequence of growth and development as other children but the speed varies.
- To help us to be concerned about the developmental stages of a child, such as sitting up, crawling and walking or so on.
- To help us to understand what should be expected from a child at each development stage.
- To provide the right environment and age appropriate resources to the children.

1.2 How Learning ability and Behaviour is Co-related to Development?

Learning means to gain knowledge, understanding and skills. An even broader term learning can be defined as any permanent change in behaviour that occurs as a result of a practice or an experience. It reveals that what children learn themselves is more important than they are taught because of its lasting affect in their behaviour.

The areas of learning and development comprise of the folowwing :-

1. Physical Development
2. Knowledge and understanding of the world
3. Communication, language and literacy
4. Personal, social and emotional development
5. Problem solving, reasoning and numeracy
6. Creative development

These six areas of learning together make up the skills, knowledge and experiences appropriate to children as they grow, learn and develop.

This paper is organized as follows. Section 2 presents some related work and recent studies on child development analysis using data mining techniques. Section 3 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 4 gives the description of the proposed approach which have shown best accuracy with our dataset. Section 5 presents the obtained results and Section 6 concludes with some remarks about the described work and guidelines for future work.

2. RELATED WORKS

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 apprach 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.

3. 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..

3.1 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.

3.2 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.

3.3 Data Selection and Transformation

The useful information is selected according to rquirements 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.

3.4 Input Variables

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

- Gross motor
- Fine motor
- Commuication
- Problem solving
- Personal, social and emotional development
- Attention and concentration
- Overactivity and impulsivity
- Passivity/ inactivity
- Planning/ organising
- Perception of space and directions
- Concepts of time
- Perception of own body
- Perception of visual forms and and figures
- Memory
- Comprehension of spoken language
- Acquisition of academic in school
- Reading, writing, arithmetic
- Social skills
- Emotional problems

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

4. PROPOSED METHOD- FUSION OF NAIVE BAYES AND DECISION TREE- 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.

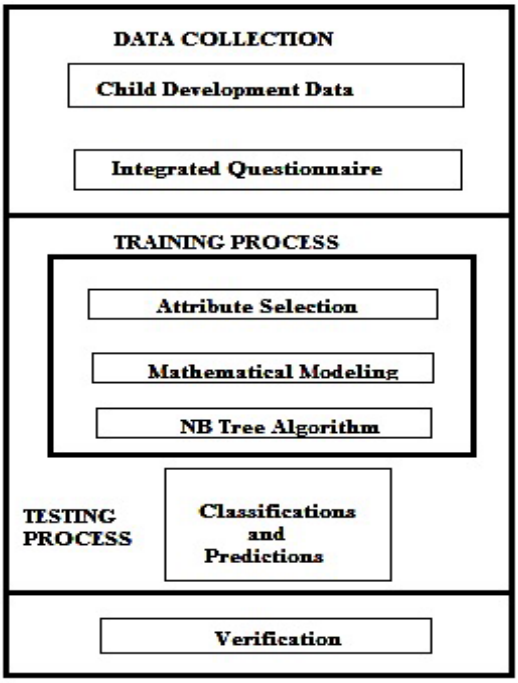

Fig -1: The prediction framework architecture

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

- Data Collections
- Child Development factors identifcation and Modelling
- Classifications and Predictions
- Verification

4.1 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.

4.2 Algorithm for Decision Tree

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 DevptFactors is empty
              then return leaf node labeled with disjunction of
          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 partitionv be elements of children set
                  with values V for property P;
                  call induce tree (partition, DevptFactors),
                  attach result to branch V
                  end
        end
end
```

4.3 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 by the ratio of its frequency in the class of interest over the prior frequency of that class.

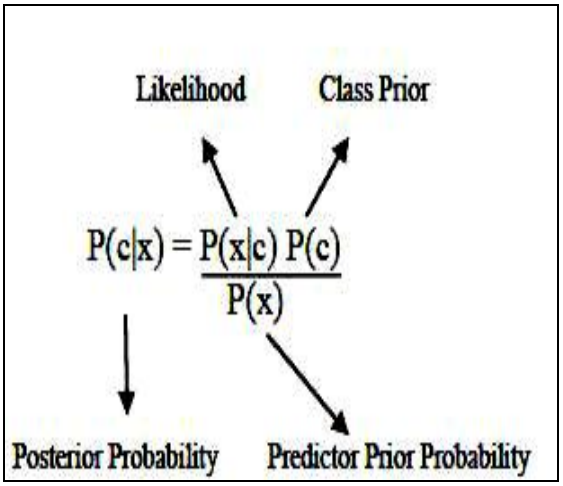

Fig -2: Naive Bayes formula

The Naive Bayes formula that we use to classify children with developmental problems are as follows:

P (x1, x2, x3....xd|C j) = P (xi|C j).....(1)

P (c|X) = P (x1|c) * P (x2|c) * P(x3|c)..... *P (xn|c) *P (c) .....(2)

For example : If a child shows defect in x1(finemotor), x2 (grossmotor), x3(communication), x4(problem solving) then the probability that a child is having a developmental defect can be calculated through the following process :-

Step 1: probability of child having poor growth can be calculated by the following method:

P (x1|C 1) = number of children having fine motor defect and have poor growth / number of children having poor growth.

P(C1) = number of children having poor growth / total number of children.

P (xn|C1) = P (x1|C1) * P (x2|C1) * P (x3|C1) * P (x4|C1) * P (C1)

Step 2: Probability of children with proper growth can be calculated as follows:

P (xn|C2) = P (x1|C2) *P (x2|C2) * P (x3|C2) * P (x4|C2) * P (C2)

Step 3: The probability of children having or not having poor growth has been compared.

If P (xn|C1) is greater then that child is having poor growth else vice-versa.

5. 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

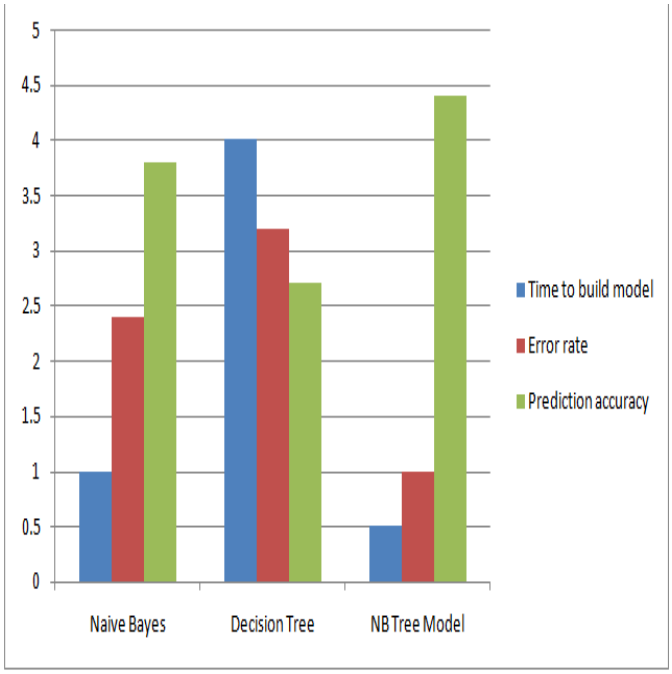

Fig -3: Comparison of the three algorithms based on Learning time, Error rate and Prediction accuracy

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.

6. CONCLUSION AND FUTURE SCOPE

In this research study, a comparitive study was conducted on various datamining classifcation 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.

In practice, NB-Trees are shown to scale to large databases and, in general, outperform Decision Trees and NBCs alone. NB-Trees appears to be a viable approach for generating prediction model especially when there are

- Many attributes are relevant for classification
- Attributes are not necessarily independent
- Database is large
- Interpretability of classifier is important

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., 1986. Induction on decision trees, Machine learning, 1(1):81-106.

[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.