Journal of Statistics Applications & Probability
*An International Journal*

http://dx.doi.org/10.18576/jsap/120337

# Modelling and Predicting Learners' Numeracy Test Results using Some Regression and Machine Learning Classifiers

*Tlhalitshi Volition Montshiwa* and Tshegofatso Botlhoko*

Department of Business Statistics & Operations Research, North West University, Mahikeng, South Africa

Received: 17 Aug. 2022, Revised: 20 Jan. 2023, Accepted: 7 Feb. 2023.
Published online: 1 Sep. 2023

**Abstract:** The prediction of early childhood numeracy skills development is often studied by determining the learner's performance in a numeracy test. It is an important study area since numeracy impacts on the learner's mathematical and statistical abilities later in life. Despite having pros and cons over each other, classification algorithms are often applied in the prediction of early childhood numeracy skills development without justifying the choice of a certain algorithm over others. In this paper, the bi-directional stepwise logistic regression model (SLRM), hierarchical logistic regression model (HLRM), classification and regression tree (CART) and Naïve Bayes (NB) were compared in terms of their ability to predict learners' numeracy test performance. The algorithms were compared using the true positive rate, true negative rate, specificity, sensitivity, classification error, classification accuracy and the area under the receiver operating characteristic curve (AUROC). The results showed that the HLRM which has been applied by several previous studies on the prediction of numeracy test competence is the best classifier followed by SLRM, CART then NB. The study also confirmed some important predictors of the learner's performance in a numeracy test some of which were also identified by some previous studies on early childhood numeracy development. Some gaps and recommendations for future research pertaining to the classification algorithms as well as implications for practice were also highlighted. We have made the HLRM scoring algorithm generated from SPSS available as a supplementary material and can be used to classify a set of new learners to either the pass or fail group.

**Keywords:** Machine learning, stepwise logistic regression, hierarchical logistic regression, classification and regression tree, naïve bayes, early childhood numeracy development.

## 1 Introduction

In many real-life applications of statistics, the choice of a statistical technique mainly depends on the objective of the research, the type of measurement scales of the data and the researcher's knowledge about the theoretical relationships between the variables. However, in the current study, we acknowledge that although the objective of the research and the measurement scales may limit the researcher to applying one statistical technique in their study, there are situations where there may be many alternative statistical methods of addressing the objectives of a particular study. In such cases, there may be a need to compare the methods and select the best approach based on some comparison criteria. In the current study, some statistical methods used by previous studies in determining the factors that impact on early childhood numeracy skills development are reviewed. We then extend the scope of these previous studies by narrowing the focus to comparing the performance of some classification and regression models in identifying the important determinants of and predicting the learners' performance on a numeracy skills test. Numeracy skills enable one to have "the knowledge and capabilities required to accommodate the mathematical demands of private and public life, and to participate in society as informed, reflective, and contributing citizens" [1]. A written numeracy test is a typical way of assessing the learner's numeracy skills; hence the interest of this paper is on determining the factors influencing the learners' performance on a numeracy test.

*Corresponding author e-mail: volition.montshiwa@nwu.ac.za

Although the paper uses data on numeracy test results, the scope of the study is more biased towards the statistical methodology around classification algorithms with the aim of contributing to this area. The current study is innovative since the classification and regression models that were compared form part of the machine learning models. In this era of the 4th industrial revolution, such machine learning models may be used in the development of systems that can automate the prediction of early childhood numeracy test results based on the variables which would be identified when training these models. This automation of the prediction of a dependent variable is artificial intelligence. As such, the present study does not only extend the scope of previous application studies by comparing some machine learning models, but its results also form the basis for identifying the determinants of numeracy test results in children and form a basis for using AI in predicting numeracy tests results. Through this study, the best classification algorithm for the prediction of numeracy test results is identified and having the best classification algorithm that produces more accurate results will enable education practitioners and planners to identify at risk learners at an early stage and to draw up some interventions in advance.

## 2 Literature review

The most commonly used classification algorithm in previous studies on early childhood numeracy skills development is the hierarchical logistic regression model (HLRM) which is also known as multilevel regression [2, 3, 4]. [5] explain that the HLRM is useful when modelling data with a group structure and a binary dependent variable where the group structure is defined by the presence of micro-observations embedded within contexts. The authors further explain that at the micro level, the usual logistic regression model (LRM) is defined for each context and then the micro coefficients are treated as functions of macro independent variables in the second step of the HLRM. Although it cannot be used for classification or prediction of early childhood numeracy performance per say, the second most commonly used method for identifying the determinants of the learners' performance on a numeracy test is the structural equation model (SEM) [6, 7, 8, 9]. Author [10] define SEM as a multivariate statistical analysis technique that is used to analyse structural relationships. The SEM as a statistical method that combines factor analysis and multiple regression analysis to determine the structural relationship between the variables. Since SEM is not a prediction or classification algorithm and cannot be directly compared to the classification algorithms considered in the current study, the SEM will not be used in the current study despite its popular use in previous studies on early childhood numeracy skills development. Other commonly used classification methods from previous studies on the prediction of learners' performance on a numeracy include the stepwise logistic regression model (SLRM) [11] and the classification and regression tree (CART) [12]. SLRM involves a step-by-step iterative process of adding or removing potential predictor variables to the model and testing for statistical significance after each iteration [13] until only the predictor variables that give the most significant model are remaining in the model. CART is a decision tree used to explain how the dependent variable can be predicted based on some iterative selection of predictor variables where each fork of the tree is a split in a predictor variable and each node at the end has a prediction for the dependent variable [14].

The models used by previous studies to predict learners' performance on a numeracy test (HLRM, SLRM and CART) fall under the umbrella of supervised machine learning classification algorithms. These algorithms are first trained using a labelled training dataset then the trained algorithm is fed on the unlabelled test dataset in order to predict the class of the dependent variable [15]. From previous studies on the prediction of the learners' performance on a numeracy test, it can be noticed that in most cases; only one of these supervised machine learning classification algorithms were used per each study and the justification for the choice of each of these models for a particular study was not usually explicitly given. The assumption made in the current study is that the choice of classification algorithms used in the already reviewed studies was based solely on the objective of these studies which was to determine the factors that affect early childhood numeracy performance and to predict this dependent variable based on the identified predictor variables. However, given that different classification algorithms were used for each study to address this same objective, the current study acknowledges that the objective of determining the factors that affect early childhood numeracy performance and its prediction can be addressed using different classification algorithms and there was a need to determine which algorithm (s) performs the best in this regard in order to justify the use of this algorithm (s) in future studies on this discipline.

This need emanates from the fact that each of the classification algorithms has its advantages and disadvantages. For example, SLRM results are dependent on the sampling error present in any sample and can lead to erroneous results; and due to its iterative and automatic selection of variables, the independent variables that it selects my not be in accordance with the theory [16]. SLRM also has a major limitation of underestimating standard errors of the parameter estimates and this leads to narrow confidence intervals which in turn leads to unreliable t-ratios [17], hence unreliable hypothesis testing results pertaining to the significance of predictor variables. [16] Further explain that the selection of independent variables for the HLRM on the other hand is not automatic; therefore, the researcher's selection of the predictor variables to be included in the HLRM is informed by theory and this gives the HLRM an advantage over the SLRM. Some advantages of the CART include good variable selection and being robust to outliers and the presence of missing values but its limitations include being sensitive to small changes in data and the likelihood of overlooking relationships between independent variables [18].

Given that they have some limitations and advantages, the need to identify the best performing classification algorithm for learners' performance on a numeracy test remain a burning issue because most previous studies in this area were application studies and did not seek to identify the most accurate algorithm for this problem. The current study seeks to bridge this gap by comparing some previously used supervised machine learning classification algorithms for predicting early childhood numeracy performance (HLRM, SLRM and CART) with the intention of identifying and recommending the best classification algorithm for future studies on the prediction of the learner's performance on a numeracy test. The current study also extends the scope of previous studies on the prediction of early childhood numeracy performance by including the naïve Bayesian (NB) algorithm, or Naïve Bayes in the comparison of the machine learning classification algorithms. [19] explains that NB is one of the most efficient and effective machine learning algorithms.

The NB algorithm is defined as a simple probability classifier through which a set of probability of an observation belonging to a certain category of the dependent variable (passing or failing the test in our case) is calculated by counting the frequency and combinations of values in a given data set [20]. [20] points out that although the assumption of conditional independence of variables under the NB methodology is rarely true in real-world applications, which makes this assumption to be naive, the NB tends to learn quickly in many classification problems. Some of the advantages of NB as identified by [21] are its ability to represent knowledge, manage complex datasets, handle small datasets and it can minimise noise in training datasets. In some previous studies, the NB has been compared to other classification algorithms in other disciplines such as in landslide susceptibility assessments against logistic regression [22], in flash flood susceptibility mapping against the Kernel Logistic Regression (KLR), Radial Basis Function Classifier (RBFC) and Logistic Model Tree (LMT) [23] and in the classification of Malaria complication against the CART [24]. However, the performance of NB against the HLRM, SLRM and CART in predicting early childhood numeracy performance remains unknown. It is in light of this gap in literature that the NB is compared to the HLRM, SLRM and CART to determine the best classification algorithm for predicting the learners' performance on a numeracy test.

## 3 Method

### *3.1 Participants*

The data used in this study is part of the Snap Survey of Ordinary Schools 1997-2016 [1] and was sourced from the Data First repository. The data comprises 13789 primary school learners aged between eight to less than thirteen years and was collected from eight provinces in South Africa. This data comprises some demographic variables of the learners, family structure variables, variables measuring the child's home learning activities and family support towards the learner's schoolwork, variables measuring some basic household resources such as the availability of water, and the dependent variable of interest which is the learner's numeracy test results. These variables are further described in Table 1.

### *3.2 Measures*

**Table 1:** Measures and variables.

| | | **PASS RATE** | | | |
|---|---|---|---|---|---|
| | | **0%-39%** (***n***=**9919, 71.9%)** | | **40%-100%** (***n***=**3870, 28.1%)** | |
| | | **Count** | **%** | **Count** | **%** |
| **Demographic variables** | | | | | |
| province | Western Cape | 731 | 7.4% | 572 | 14.8% |
| | Kwa-Zulu Natal | 2036 | 20.5% | 796 | 20.6% |

| | | | | | |
|---|---|---|---|---|---|
| | North-West | 806 | 8.1% | 308 | 8.0% |
| | Eastern Cape | 1510 | 15.2% | 662 | 17.1% |
| | Northern Cape | 722 | 7.3% | 246 | 6.4% |
| | Mpumalanga | 1556 | 15.7% | 634 | 16.4% |
| | Limpopo | 1675 | 16.9% | 284 | 7.3% |
| | Free-State Province | 883 | 8.9% | 368 | 9.5% |
| gender | MALE | 5154 | 52.2% | 1881 | 48.7% |
| | FEMALE | 4727 | 47.8% | 1983 | 51.3% |
| actual age | 8-<10 YEARS | 8306 | 83.7% | 3460 | 89.4% |
| | 11 -<13 YEARS | 1613 | 16.3% | 410 | 10.6% |
| home language | ISIXHOSA | 1780 | 17.9% | 685 | 17.7% |
| | AFRIKAANS | 913 | 9.2% | 822 | 21.2% |
| | SESOTHO | 956 | 9.6% | 225 | 5.8% |
| | SETSWANA | 1104 | 11.1% | 268 | 6.9% |
| | ISIZULU | 2347 | 23.7% | 804 | 20.8% |
| | ISINDEBELE | 268 | 2.7% | 72 | 1.9% |
| | ENGLISH | 148 | 1.5% | 437 | 11.3% |
| | SISWATI | 393 | 4.0% | 185 | 4.8% |
| | tshiVENDA | 210 | 2.1% | 27 | 0.7% |
| | XiTSONGA | 561 | 5.7% | 84 | 2.2% |
| | SEPEDI | 1239 | 12.5% | 261 | 6.7% |
| **Family structure** | | | | | |
| stay with mother | NO | 2080 | 21.0% | 641 | 16.6% |
| | YES | 7839 | 79.0% | 3229 | 83.4% |

| | | | | | |
|---|---|---|---|---|---|
| stay with father | NO | 4662 | 47.0% | 1729 | 44.7% |
| | YES | 5257 | 53.0% | 2141 | 55.3% |
| stay with grand mother | NO | 5361 | 54.0% | 2426 | 62.7% |
| | YES | 4558 | 46.0% | 1444 | 37.3% |
| stay with grandfather | NO | 7310 | 73.7% | 3138 | 81.1% |
| | YES | 2609 | 26.3% | 732 | 18.9% |
| stay with aunt | NO | 6646 | 67.0% | 2881 | 74.4% |
| | YES | 3273 | 33.0% | 989 | 25.6% |
| stay with uncle | NO | 7281 | 73.4% | 3046 | 78.7% |
| | YES | 2638 | 26.6% | 824 | 21.3% |
| other children | BETWEEN 1 AND 4 | 6978 | 70.3% | 3026 | 78.2% |
| | MORE THAN 5 CHILDREN | 2941 | 29.7% | 844 | 21.8% |
| **At home learning activities and family support** | | | | | |
| adult reads stories at home | NEVER | 4043 | 40.8% | 1773 | 45.8% |
| | MORE THAN 2 TIMES A WEEK | 5876 | 59.2% | 2097 | 54.2% |
| how often adult read to child | NEVER | 6544 | 66.0% | 2571 | 66.4% |
| | MORE THAN 2 TIMES A WEEK | 3375 | 34.0% | 1299 | 33.6% |
| how often read alone | NEVER | 4876 | 49.2% | 1468 | 37.9% |
| | MORE THAN 2 TIMES A WEEK | 5043 | 50.8% | 2402 | 62.1% |
| how often do child do homework | NEVER | 3929 | 39.6% | 964 | 24.9% |

| | | | | | |
|---|---|---|---|---|---|
| | MORE THAN 2 TIMES A WEEK | 5990 | 60.4% | 2906 | 75.1% |
| does adult help with homework | NEVER | 1379 | 13.9% | 441 | 11.4% |
| | MORE THAN 2 TIMES A WEEK | 8540 | 86.1% | 3429 | 88.6% |
| mother help | NEVER | 5110 | 51.5% | 1735 | 44.8% |
| | MORE THAN 2 TIMES A WEEK | 4809 | 48.5% | 2135 | 55.2% |
| father help | NEVER | 8526 | 86.0% | 3071 | 79.4% |
| | MORE THAN 2 TIMES A WEEK | 1393 | 14.0% | 799 | 20.6% |
| sister help | NEVER | 6390 | 64.4% | 2727 | 70.5% |
| | MORE THAN 2 TIMES A WEEK | 3529 | 35.6% | 1143 | 29.5% |
| brother help | NEVER | 8333 | 84.0% | 3226 | 83.4% |
| | MORE THAN 2 TIMES A WEEK | 1586 | 16.0% | 644 | 16.6% |
| **Basic household resources and resources** | | | | | |
| electricity | NO | 3224 | 32.5% | 891 | 23.0% |
| | YES | 6695 | 67.5% | 2979 | 77.0% |
| tap water | NO | 5039 | 50.8% | 1362 | 35.2% |
| | YES | 4880 | 49.2% | 2508 | 64.8% |
| toilet in house | NO | 6867 | 69.2% | 1908 | 49.3% |
| | YES | 3052 | 30.8% | 1962 | 50.7% |
| car | NO | 6160 | 62.1% | 1769 | 45.7% |

| | | | | | |
|---|---|---|---|---|---|
| | YES | 3759 | 37.9% | 2101 | 54.3% |
| computer | NO | 8351 | 84.2% | 2575 | 66.5% |
| | YES | 1568 | 15.8% | 1295 | 33.5% |
| newspaper everyday | NO | 5572 | 56.2% | 1891 | 48.9% |
| | YES | 4347 | 43.8% | 1979 | 51.1% |
| fridge | NO | 3969 | 40.0% | 986 | 25.5% |
| | YES | 5950 | 60.0% | 2884 | 74.5% |
| washing machine | NO | 7640 | 77.0% | 2104 | 54.4% |
| | YES | 2279 | 23.0% | 1766 | 45.6% |

### *3.2 .1 Family structure*

The family structure variables are all binary (0=No and 1=Yes) and they determine whether the learner stays with their mother, father, grandmother, grandfather, aunt, uncle and whether there are other children in the learner's household. The previous studies reviewed in this current study [2, 3, 4, 5, 6, 7, 8, 9, 10] did not include the presence of or absence of these family members in their models. As such, the inclusion of these family structure variables extends the scope of these previous studies. Family structure variables are important to explore. For example, [11] found that living in single-mother families has a negative effect on the time that the child invests in studying, reading and doing homework. In another study, [12] found that the extent to which mothers and fathers equally share responsibilities for playful activities may promote the child's cognitive development. The other study that highlights the impact of family structure on early childhood development is the one conducted by [13] which established that having an older sibling in general is associated with increased working memory whereas having an older sister is related to increased working memory and cognitive flexibility in young children. As such, the family structure variables' impact on the learner's performance in a numeracy test and they will be tested in the current study alongside the other variables which are discussed below.

### *3.2.2 At home learning activities and family support*

Author [14] who conducted a study to determine the role played by home numeracy related activities on the development of early numeracy skills explain that parents play a very important role in the development of early numeracy. As such, the current sought to extend this argument by determining whether the number of times in which the parents assist the child with homework has an impact on the development of the performance in a numeracy test. The study extends this enquiry to determine whether the support from other adults in the household such as the learner's siblings also has an impact on the development of the child's numeracy skills.

The rationale behind the inclusion of family support variables also arises from previous studies such as the ones conducted by [8] who determined how family and individual variables are associated with the numeracy interest and competence. [7]found that children's numeracy interest was linked with their parents' practices and attitudes. [4] found that mothers' engagement in numeracy practices at home had an impact on the children's numeracy performance. [2] also found that Home learning environment (HLE) which was determined by the extent to which primary caregiver and the child do activities related to numeracy is associated with numeracy skills in the first year of preschool. [3] established that parent-child collaborative activities increased the child's exposure to numeracy activities. It was therefore important to include family support variables in the algorithms that were evaluated in the current study.

### *3.2.3 Basic household infrastructure and resources*

In their study, [15] explained that there is evidence that the extent to which the child is experiences poverty (persistent, transient or non-poor) affect the child's cognitive and behavioural traits. The availability of resources such as a car, fridge and washing machine which eases the day-to-day chores of the learner, and their family members are some of the variables which are important to consider when defining poverty. As such, the effect of these variables on the child's numeracy skills development is also of interest to the current study. Access to tap water and electricity in one's household are also some of the variables which are important to consider when defining poverty. As such, the effect of these variables on the leaners' performance on a numeracy test is also of interest to the current study. Certain previous studies have also tested the impact of these variables on the child's performance in a numeracy test. For example, [5] included a variable measuring water and sanitation facilities in their model. Several previous studies have also established that the socio-economic status of the learner has an impact on the learner's numeracy skills attainment.

### *3.2.4 Demographic variables of the learners*

The child's demographic variables (province of origin, gender, age and home language) will also be included in the classification algorithms under study. Several studies have established that these variables have an impact on the child's performance pertaining to numeracy skills test. [5] found that height-for-age has a positive impact on the child's literacy–numeracy and learning development; [7] found that children's numeracy competence is related to their gender and age, and [2] also established that gender and native language status impacted the children's numeracy competence.

### *3.2.5 Learner's numeracy test results*

The dependent variable in the current study is the learner's numeracy test results. The pass mark ranges from 40% to 100% and 0% to 39% is considered to be a fail. The dependent variable is a binary variable in which 0 denotes a fail and 1 denotes a pass. The learners who failed the test comprise 71.9% ($n$=9919) and those who passed the numeracy test comprise 28.1% ($n$=3870) of the whole sample.

## 4 Data analysis procedure

Four supervised machine learning classification algorithms namely the hierarchical logistic regression model (HLRM), stepwise logistic regression model (SLRM), classification and regression tree (CART) and naïve Bayes (NB) are compared on the basis of their ability to classify learners according to their competence in a numeracy test using family structure, family support, basic household infrastructure, basic household resources, demographic variables of the learners and learner's numeracy test results. In this section, the mathematical formulations of these learning algorithms are briefly introduced, and the criteria used in comparing these algorithms are also discussed.

### ***SLRM***

The SRLM was first proposed by Efroymson in 1960 and it is an iterative method for choosing predictor variables for a regression model [16]. There are three variants of stepwise regression which are forward selection, backward selection, and bi-directional selection. At each step of the **forward selection** method, the predictor variable which that gives the most significant value for the new model from all candidate predictors for entry is added to the model [17]. For the **backward selection** approach, the process starts with all predictor variables included in the model and at each step, a predictor variable is removed so that the variables remaining yield the most significant new model that is better than the previous model. For the **bi-directional selection approach**, the forward selection method is used to select the variables that are eligible for entry, and the backward selection method is used to remove the variables that were eligible for exclusion from the model based on the SBC. Since it encompasses both the forward and backward selection, the bi-directional stepwise regression was the one that was applied in the current study and its results were compared to that of the other three algorithms under study. The iteration selection and/or removal of variables for the bi-directional approaches were set to be terminated when the SBC for the validation dataset is greater than that of the training dataset at a particular step [17]. All data analyses for the SLRM were performed using the HPLOGISTIC procedure in Statistical Analysis System (SAS) version 9.4.

After selecting the predictor variables, the model for predicting the learner's numeracy test results was automatically fitted using the conventional multivariate binary logistic regression model using only the variables selected by the bi-directional selection approach. Through the conventional multivariate binary logistic regression model, the odds of a learner passing the numeracy test were compared to the odds of not passing the test based on the odds ratio (OR). The OR is represented by Equation 1 [18]:

$$OR = \frac{p}{1-p}, \tag{1}$$

where $p$ is the probability of a learner passing the numeracy test whereas 1-$p$ is the probability of a learner failing the numeracy test.

The logistic regression analyses the natural logarithmic (Ln) transformation of the odds which is known as the log odds or logit. The logit is computed using Equation 2 [18]:

$$Logit(p) = \ln\left[\frac{p}{1-p}\right] = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_i X_i, \tag{2}$$

where $\hat{\beta}_i$'s for $i=0...I$ are estimates of the coefficients of the predictor variables and $X_i$'s are the predictor variables remaining in the model after stepwise selection was implemented. Equation 2 can also be expressed as the probability of the learner passing the numeracy test ($p(y=1)$) given a set of predictor variables that were selected by the stepwise methods ($X_i$'s) which will yield Equation 3 [18]:

$$p(y = 1 \mid X_i\text{'s}) = \frac{1}{\mathbf{1}+\boldsymbol{e}^{-(\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2+\cdots+\hat{\beta}_i X_i)}} \tag{3}$$

The parameters such as the coefficients of the model together with their standard errors and p-values are estimated by maximising the likelihood of the model. The likelihood in our case is the probability that the observed values of the numeracy test results may be predicted from the observed values of the stepwise selected predictor variables remaining in the final models. This method of estimating parameters for a regression model is known as the maximum likelihood approach.

## *CART*

Author [19] explain that the CART was developed by Breiman, Friedman, Olshen and Stone in 1984, and [20] describe it as a non-parametric data mining technique that is used to create a decision tree that in turn can be used to classify observations based on a given outcome. A decision tree is a classification method that uses multiple predictor variables to create prediction algorithms for a dependent variable and it is represented by a diagram which depicts a tree with a root node, internal nodes, and leaf nodes [21]. The root node or decision node marks the beginning of the tree, and it divides all observations into two or more mutually exclusive subsets, internal nodes or chance nodes is a representation of one of the possible choices available at a particular point in the tree and the leaf nodes or end nodes represents the result of a combination of decisions [21]. The HPSPLIT procedure was used to fit the CART in SAS 9.4 where Entropy was used for growing the tree. Entropy is one of the nodes splitting criteria whose aim is to reduce the impurity of a node [22]. For a dataset $\boldsymbol{D}$, the Entropy impurity measure for a binary dependent variable is computed using Equation 4 [23]:

$$Entropy(\boldsymbol{D}) = \sum_{i=1}^{g} -p_i \log_2 p_i, \tag{4}$$

where $p_i$ is the probability that an arbitrary tuple in dataset $\boldsymbol{D}$ belongs to class $g$.

At each node, CART aims to maximise the information gain which is the difference between the original information and the information obtained after the partitioning the node [23] and it is also known as the goodness of split [24]. At each step, the information gain is computed using Equation 5 [23]:

$$Gain(\boldsymbol{D},\boldsymbol{X}) = Entropy(\boldsymbol{D}) - \sum_{j=1}^{v} \frac{|\boldsymbol{D}_J|}{|\boldsymbol{D}|} Entropy(\boldsymbol{D}_j), \tag{5}$$

where $X$ is a predictor variable at a particular node and $\boldsymbol{D}_j$ are elements of a subset of partitions of $\boldsymbol{D}$**.** The node is split only if splitting leads to an improvement in the information, otherwise it is not split. It is worth noting that in many practical uses of CART, not all candidate predictor variables are used in the tree and it is possible that one variable can be used in more than one node [21].

Over-fitting in CART creates a complex model which is likely to yield unreliable results when such a model is used for prediction, and such complex models are not generalisable and they lack robustness [21]. In the current study, post pruning was used after creating the tree to optimise the computational efficiency and classification accuracy of the tree [25]. Post-pruning also reduces the unnecessary complexity of the tree by reducing its size and this will lead to a less cumbersome structure of the tree [25] and in turn will make the tree more useful for prediction. For the current study, the Quinlan's 1987 reduced-error pruning was used. The method performs pruning by starting with a full tree $T_0$, then it creates a subtree by replacing a node from the full tree with a leaf and calculates the error rate from the validation dataset [26]. The node with the smallest error rate is then replaced with a leaf and a subtree $T_1$ is created. The process is repeated and the subtree that has the smallest error rate is selected as the final subtree [26].

### *NB*

Naïve Bayes classification algorithm is based on the Bayes Theorem which was first discovered by Thomas Bayes in 1763 [27]. The algorithm assumes that each predictor variable makes an independent and equal contribution to the dependent variable [28]. The Bayes Theorem states that [29]:

$$p(G_j|\boldsymbol{x}) = \frac{p(G_j)p(x|G_j)}{p(x)}, \tag{6}$$

where $G_j$ are groups of the dependent variable, $\boldsymbol{x}$ is the vector of independent variables, $p(G_j|\boldsymbol{x})$ is the posterior conditional probability of belonging to a group $G_j$ of the dependent variable given the vector of predictor variables **x**, $p(G_j)$ is the prior probability of belong to a group of the dependent variable, $p(\boldsymbol{x}|G_j)$ is the likelihood which is the probability of the independent variable given a group of the dependent variable, and $p(\boldsymbol{x})$ is prior probability of the independent variable.

By letting the vector of *n* independent variables in the current study to be $[x_1, x_2, \dots, x_n]$ then Equation 6 can be re-written as:

$$p(G_j|x_1, x_2, \dots, x_n) = \frac{p(G_j)p(x_1|G_j)p(x_2|G_j)\dots p(x_n|G_j)}{p(x_1)p(x_2)\dots p(x_n)} \tag{7}$$

Author [28] explain that for a specific dataset, the denominator of Equation 7 is constant, therefore; the equation can be written as the following simplified proportionality:

$$p(G_j|x_1, x_2, \dots, x_n) \propto p(G_j)p(x_1|G_j)p(x_2|G_j)\dots p(x_n|G_j) \tag{8}$$

The Naïve Bayes classification algorithm generates a label $G$ which estimates the probability of a given sample with known values of the independent variables to belong to a certain group of the dependent variable by picking the $G_j$ that maximises $p(G_j)p(x_1|G_j)p(x_2|G_j)\dots p(x_n|G_j)$ . That is,

$$G = argmax_{G_j}\, p(G_j)p(x_1|G_j)p(x_2|G_j)\dots p(x_n|G_j) \tag{9}$$

In the current study, the Naïve Bayes algorithm is fitted using GaussianNB from the Scikit-learn module of the Python 3.9 software. The parameter estimates for Naïve Bayes were estimated using maximum likelihood.

### *HLRM*

The HLRM is used to analyse nested sources of variability by representing clustering impacts of observations (in the current study these are the learners) within groups of higher-level units when assessing the impact of predictor variables on the dependent variable (in this study this is the learner's results from a numeracy test results) [44, 45]. In other words, the impact of predictor variables on the variability in the dependent variable is determined while taking the variability associated with each cluster/ hierarchy into account [44]. This approach is also referred to as multi-level regression. In this study, the HLRM was fitted using the Statistical Package for Social Sciences (SPSS) through the generalised linear mixed models module. The first step was to determine whether the candidate random effects for the model known as level 2 variables and/ or their interaction effects are significant before they were entered in the model. This was achieved by testing whether the intercept of the HLRM randomly varies across the level 2 variables. After identifying significant random effects, the HLRM was fitted by following two main steps.

In the basic step of the HLRM, the learner's results from a numeracy test were predicted as a linear combination of level 1 variables plus the intercept. The level 1 variables in the current study are family structure variables, at home learning activities and family support variables, and basic household resources. The level 1 model of the HLRMs is explained by Equation 9 [30]:

$$Y_{ij} = \alpha_{0i} + \alpha_{1j}X_1 + \cdots + \alpha_{kj}X_k + r_{ij}, \quad (10)$$

where $\alpha_{0i}$ denotes the intercept of group *j*, $\alpha_{1j}$ denotes the slopes for the independent variable $X_1$ of group *j* and $r_{ij}$ is the residual for learner *i* within group *j*.

In the second step of HLRM, the level 1 parameter estimates (slopes) and intercept from the level 1 model are treated as dependent variables which are now predicted from level 2 variables. In the current study, the candidate level 2 variables were demographic variables which are province, gender, actual age and home language. That is, the relationship between level 1 variables and the learner's numeracy test results is determined while assuming that the learners are nested within their demographic variables or the interaction between some of them. Some of these demographic variables may not have any clustering effect and will be included as level 1 variables. Equations 11 and 12 represent the first two forms of level 2 modelling in HLRM [30]:

$$\alpha_{0i} = \gamma_{00} + \gamma_{01}V_1 + \cdots + \gamma_{0k}V_k + \varepsilon_{0i}, \quad (11)$$
$$\alpha_{1i} = \gamma_{10} + \gamma_{11}V_1 + \cdots + \gamma_{1k}V_k + \varepsilon_{0i} \quad (12)$$

and so on. From Equations 11 and 12, $\gamma_{00}$ and $\gamma_{10}$ are intercepts, $\gamma_{01}$ and $\gamma_{11}$ are slopes predicting $\alpha_{0i}$ and $\alpha_{1i}$ respectively from variable $V_1$. The HLRM process assists one to accurately model the effects of level 1 and level 2 variables on the learner's numeracy test results, and it can aid in determining the cross-level interactions to understand the determinants of the differences in the relationship between level 1 variables and the learner's numeracy test results. The parameters of the model were estimated using maximum likelihood.

## *Testing and Validation*

The classifications algorithms used in the current study were first trained and the trained algorithm was then validated in order to assess their predictive ability. As such, the data in this study was split into 70% training data and 30% validation data which is a commonly used training- to- validation data splitting ratio.

## *Model Comparison Criteria*

The classification algorithms were compared using classification accuracy, classification error, precision, specificity, sensitivity (also known as recall) and the area under the receiver operating characteristic (AUROC) curve. These approaches have been used by many previous studies on the comparison of the effectiveness of machine learning models such as the studies by [31, 32]. The classification accuracy, classification error, precision, specificity, and sensitivity are computed from the confusion matrix which shows the number of correctly classified events (true positives) and non-events (true negatives), and incorrectly classified events (false negatives) and non-events (false positives) from the fitted classification algorithm using the validation dataset. The general confusion matrix is represented in Table 2.

**Table 2**: Confusion Matrix.

| | | **Predicted** | |
|---|---|---|---|
| | | **40%-100%** | **0%-39%** |
| **Actual** | **40%-100%** | True Positive (TP) | False Negative (FN) |
| | **0%-39%** | False Positive (FP) | True Negative (TN) |

Using the statistics from Table 2, the following model comparison criteria are further calculated:

$$Classification\ accuray = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

$$Classification\ error = \frac{FP+FN}{TP+TN+FP+FN} = 1 - Classification\ accuracy \quad (14)$$

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

$$Specificity = \frac{TN}{TN+FP} \quad (16)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (17)$$

The statistics from Table 2 can also be reported as rates by dividing each one of them by the size of the validation dataset and multiplying the answer by 100%. The same can also be done for the statistics that are founded from Equations 13 through to 17.

In addition, the algorithms in the current study will also be compared using the AUROC which is the probability that a randomly chosen observation from the 0%-39% group will have a smaller estimated probability of belonging to the 40%-100% group than a randomly chosen observation from 40%-100% group. The AUROC is computed using Equation 18 [33]:

$$AUROC = \frac{\sum r_i - p(p+1)/2}{pn}, \quad (18)$$

where $\sum r_i$ is the rank of the $i^{\text{th}}$ observation from the 40%-100% group, and $p$ and $n$ are the total numbers of observations from the 40%-100% and the 0%-39% groups respectively. The algorithm with the highest classification accuracy, precision, specificity, sensitivity and AUROC as well as the lowest classification error was preferred.

## 5 Results

**Table 3:** Confusion matrices for classification algorithms.

| | | | Predicted | |
|---|---|---|---|---|
| **SLRM** | | | **40%-100%** | **0%-39%** |
| | **Actual** | **40%-100%** | (8%) | (21%) |
| | | **0%-39%** | (3%) | (68%) |
| | | | **Predicted** | |
| **Naive Bayes** | | | **40%-100%** | **0%-39%** |
| | **Actual** | **40%-100%** | (59%) | (14%) |
| | | **0%-39%** | (14%) | (13%) |
| | | | **Predicted** | |
| **CART** | | | **40%-100%** | **0%-39%** |
| | **Actual** | **40%-100%** | (68%) | (4%) |
| | | **0%-39%** | (19%) | (9%) |
| | | | **Predicted** | |
| **HLRM** | | | **40%-100%** | **0%-39%** |
| | **Actual** | **40%-100%** | (5%) | (17%) |

| | | | | |
|---|---|---|---|---|
| | | **0%-39%** | (2%) | (76%) |

Table 3 shows that when ranked from highest to lowest, the CART, Naive Bayes, SLRM and HLRM correctly classify 68%, 59%, 8% and 5% of learners who passed the numeracy test (40%-100%) respectively. This implies that CART is the best algorithm in terms of the true positives rate. The table also shows that when ranked from highest to lowest, the HLRM, SLRM, Naive Bayes and CART correctly classify 76%, 68%, 13% and 9% of learners who failed the numeracy test (40%-100%) respectively. This implies that HLRM is the best algorithm in terms of the true negatives rate. The algorithm with the worst misclassification of the learners who passed the test is the SLRM with 21% of the learners being misclassified as having failed the test, followed by HLRM (17%), then Naive Bayes (14%), but only 4% of the learners who passed the test were misclassified as having failed the test by CART. As such, CART is the best algorithm in terms of minimising the misclassification of learners who passed the test. In addition, the algorithm with the worst misclassification of the earners who failed the test is the CART with 19% of the learners being misclassified as having passed the test, followed by Naive Bayes (14%), the third worst algorithm relative to misclassification of learners who failed the test is SLRM (3%), but only 2% of the learners who failed the test were misclassified as having passed the test by HLRM. As such, HLRM is the best algorithm in terms of minimising the misclassification of learners who failed the test.

**Table 4:** Classification algorithms comparison criteria.

| **Classification Algorithm** | **Classification accuracy** | **Classification error** | **Specificity** | **Sensitivity (recall)** | **Precision** | **AUROC** |
|---|---|---|---|---|---|---|
| SLRM (Bi-directional) | 0.7518 | 0.2482 | 0.9607 | 0.2577 | 0.7350 | 0.7122 |
| HLRM | 0.8108 | 0.1892 | 0.9783 | 0.2884 | 0.7297 | 0.5943 |
| CART | 0.7691 | 0.2309 | 0.9420 | 0.2816 | 0.6718 | 0.7003 |
| NB | 0.7177 | 0.2823 | 0.4768 | 0.8099 | 0.8019 | 0.6433 |

Following the computation of classification accuracy, classification error, specificity, sensitivity, precision and AUROC, the algorithms were ranked based on their overall classification ability and the ranking of the algorithms is detailed in Table 5.

**Table 5:** Ranking of the classification algorithms.

| **Classification Algorithm** | **Classification accuracy (Rank)** | **Classification error (Rank)** | **Specificity (Rank)** | **Sensitivity (recall)** | **Precision (Recall)** | **AUROC** | **Overall Rank (The lower the better)** |
|---|---|---|---|---|---|---|---|
| SLRM | 3 | 3 | 2 | 4 | 2 | 1 | 15 |
| HLRM | 1 | 1 | 1 | 2 | 3 | 4 | 12 |
| CART | 2 | 2 | 3 | 3 | 4 | 2 | 16 |
| NB | 4 | 4 | 4 | 1 | 1 | 3 | 17 |

The classification accuracy measure in Table 4 shows that out of all the learners in the validation dataset, the algorithms classified more than 70% with HLRM giving the best classification accuracy (81.08%) whereas Naive Bayes gave the worst classification accuracy (71.77%). The classification error is the percentage of observations remaining after the correctly classifying some observations (1-classification accuracy), therefore; it is obvious that the HLRM gave the lowest classification error (18.92%) whereas Naive Bayes gave the highest error (28.23%). The table also shows that except for Naive Bayes which has a specificity rate of 47.68%, all other algorithms have a sensitivity of more than 90% and HLRM has the highest specificity (97.83). This implies that, Naive Bayes severely misclassified the passes, whereas HLRM gave the best classification of the passes. On the other hand, Table 4 shows that Naive Bayes gives the best classification of the occurrences of failure (80.99%) whereas the other algorithms gave a poor classification of the occurrences of failure (at least 25%) whereby SLRM severely misclassified of the occurrences of failure (25.77%). In terms of precision, the Naive Bayes gave the best results (80.19%) whereas CART yielded the worst results (67.18%). However, in terms of AUROC, SLRM gave the best results (71.22%), but Naive Bayes gave the lowest AUROC (64.33%). The comparison of the models using the statistics from Table 4 is further depicted in Figure 1.

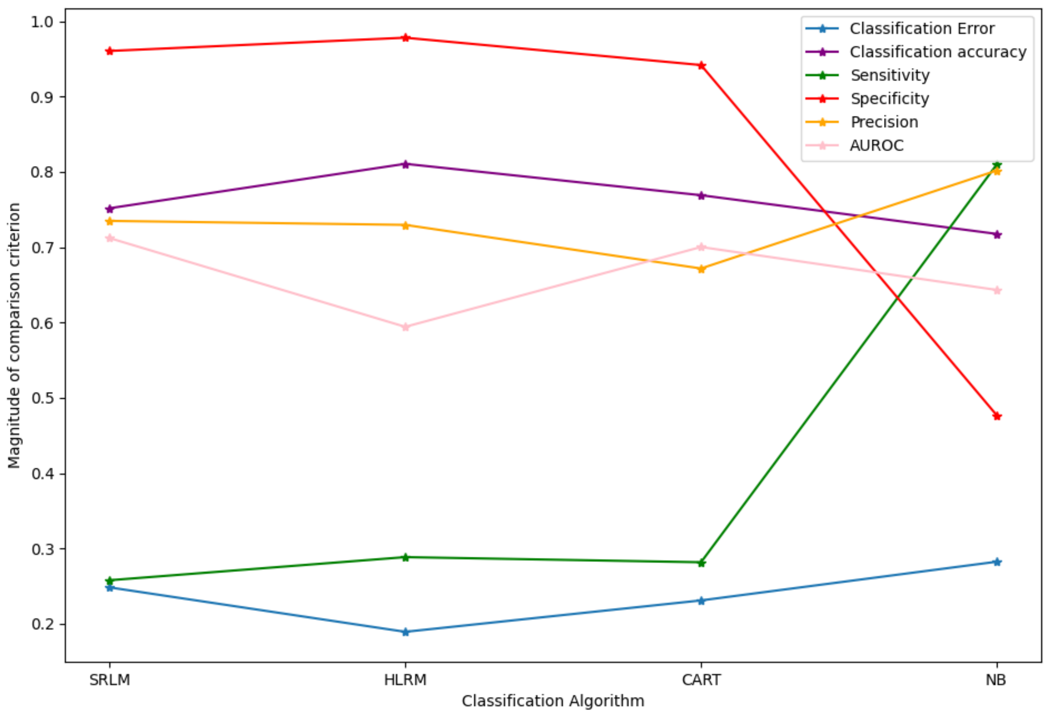


**Fig. 1:** Comparison of Classification algorithms.

The ranking of the models in Table 5 is based on the model comparison criteria which are computed from the confusion matrices from Table 3 and are displayed in Figure 1. Under each model comparison criterion in Table 5, the model with the lowest rank is the best performer. That is, the models are ranked from good (1) to worse (4) in terms of their classification ability. As such, HLRM is the best algorithm for classifying learners relative to their performance on a numeracy test based on the predictor variables that were considered in the current study. The second-best algorithm is SLRM, followed by CART and NB is the least performing algorithm in terms of classification of these learners. The parameter estimates for the chosen model (HLRM) are discussed next.

**Table 6:** Parameter estimates for the random effect.

| Random Effect Covariance | Estimate | Std. Error | Z | p-value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Var(Intercept) | .896 | .224 | 3.997 | .000 | .548 | 1.462 |

| |
|---|
| Covariance Structure: Variance components |
| Subject Specification: homelanguage * province |

Only the interaction between home language and province had a significant random effect with a p-value of 0.000 as shown in Table 6. This implies that the relationship between level 1 variables (predictor variables) and the learner's performance on a numeracy test varies randomly across the interaction effect between home language and province. As such, the regression model to determine impact of predictor variables on the variability in the learner's performance on a numeracy test was then determined while taking the variability associated with the interaction effect between language and province into account and the parameter estimates for this model are presented in Table 7.

**Table 7:** Parameter estimates for the fixed effects.

| **Model Term** | $\hat{\beta}$ | **Std. Error** | **t** | **p-value** | **Exp($\hat{\beta}$)** |
|---|---|---|---|---|---|
| Intercept | -1.397 | .1678 | -8.325 | .000 | .247 |
| gender=Male | .048 | .0489 | .990 | .322 | 1.050 |
| actual age=1 | -.366 | .0737 | -4.962 | .000 | .694 |
| stay with mother=1 | .202 | .0709 | 2.847 | .004 | 1.224 |
| stay with father=1 | -.262 | .0550 | -4.771 | .000 | .769 |
| stay with grandmother=1 | -.097 | .0577 | -1.685 | .092 | .907 |
| stay with grandfather=1 | -.232 | .0664 | -3.488 | .000 | .793 |
| stay with aunt=1 | -.143 | .0614 | -2.332 | .020 | .867 |
| stay with uncle=1 | -.006 | .0657 | -.097 | .922 | .994 |
| other children=1 | -.158 | .0569 | -2.777 | .005 | .854 |
| adult reads stories at home=1 | -.216 | .0605 | -3.565 | .000 | .806 |
| how often adult read tochild=1 | -.048 | .0644 | -.741 | .459 | .953 |
| how often read alone=1 | .257 | .0544 | 4.731 | .000 | 1.294 |
| how often do child do homework=1 | .271 | .0568 | 4.778 | .000 | 1.312 |
| does adult help with homework=1 | .158 | .0800 | 1.979 | .048 | 1.171 |
| mother help=1 | -.098 | .0557 | -1.753 | .080 | .907 |
| fatherhelp=1 | .036 | .0698 | .512 | .609 | 1.036 |
| sister help=1 | -.142 | .0546 | -2.608 | .009 | .867 |
| brother help=1 | .038 | .0663 | .579 | .562 | 1.039 |
| electricity=1 | -.021 | .0647 | -.323 | .747 | .979 |
| tap water=1 | .048 | .0651 | .741 | .459 | 1.049 |
| toilet in house=1 | .070 | .0652 | 1.067 | .286 | 1.072 |

| | | | | | |
|---|---|---|---|---|---|
| car=1 | .238 | .0544 | 4.377 | .000 | 1.269 |
| computer=1 | .197 | .0682 | 2.894 | .004 | 1.218 |
| newspaper everyday=1 | .004 | .0546 | .071 | .943 | 1.004 |
| fridge=1 | .311 | .0638 | 4.870 | .000 | 1.364 |
| washing machine=1 | .278 | .0704 | 3.944 | .000 | 1.320 |

From the results shown in Table 7, only the effect of variables with significant p-values of less than 0.05 are interpreted and insignificant predictor variables are not considered in this section. Exp ($\hat{\beta}$) are the odds ratios and they explain the effect of a significant predictor on the odds or chance of a learner passing the numeracy test.

From Table 7, it can be noticed that gender and actual age were the only demographic variables which were included as predictors in the model since they did not have a significant cluster effect on the variation of pass rate unlike home language and province which are not included in this table. However, only actual age has a significant effect on the pass rate (p-value =0.000) whereby the odds of passing a numeracy test for learners in category 1 of the age group (11 to <13 years) are 0.357 (1-0.643) times less than that of learners in the younger age group (8- 10 years).

From the *family structure* variables in Table 7, the odds of passing the numeracy test for learners who stay with their mothers are 0.224 (1.224-1) higher than those who do not stay with their mothers. However, having a learner staying with fathers, with their grandfathers, with their aunts or with other children in their households decreases the odds of passing the numeracy test by 0.231(1-0.769), 0.207(1-0.793), 0.133 (1-0.867) and 0.146 (1-0.854) respectively.

For the *at home learning activities and family support* in Table 7, the odds of passing a numeracy test by learners who often read alone more than two times a week are 0.294 (1.294-1) times more than for those who never read alone at home. Also, the odds of passing the numeracy test by learners who do more homework more than two times a week are 0.312 (1.312-1) times more than for those who never do homework, and the odds for learners who get help with homework from an adult more than 2 times a week are 0.171 (1.171-1) times more than for those who never gets help. On the other hand, getting help from a sister while doing school related activities at home decreases the odds of passing a numeracy test by 0.133 (1-0.867) compared to those who never get help from their sisters.

From *basic household infrastructure and resource* in Table 7, the odds of passing the numeracy test by learners who have a car, computer, fridge or a washing machine are 0.269 (1.269-0.269), 0.218 (1.218-1), 0.364 (1.364-1) and 0.320 (1.320-1) times more than for learners who do not have these resources at home. All other potential predictor variables of the learner's performance were found to be insignificant in the HLRM.2

## 6 Contribution to practice

We have made the HLRM scoring algorithm generated from SPSS available as a supplementary material. This algorithm can be used to classify a set of new learners to either the pass or fail group. The algorithm only works with data that contains the variables names (measured on a binary scale) that were used in our study. When making recommendations using the predicted classifications from the algorithm, users should exercise caution and pay attention to the limitations of this model as noted in this paper. Users can follow these simple steps to use this algorithm: Open the dataset in SPSS, choose the UTILITIES tab, then choose the SCORING WIZARD, Import the HLRM scoring algorithm, click next twice, then set the value of the Probability of Selected Category to one (i.e. pass) and then click on FINISH. The predicted category from the new dataset will be saved as a variable in that dataset.

## 7 Conclusions

Previous studies reviewed in this study were focused on the application of some classification algorithms in the prediction of the learner's performance in early childhood numeracy development without justification of the choice of the model or algorithm used in such studies. However, each classification algorithm has its own advantages and disadvantages, and the classification ability of these may vary from one area of application to the other so this study sought to determine the best classification algorithm for the data under study. Based on the results of the study, the four algorithms either give a good classification of learners who passed the test while severely misclassifying the ones who failed the test (CART and Naive Bayes) or the algorithms give a good classification of learners who failed the test while severely misclassifying the ones who passed the test (HLRM and SLRM). However, after a further comparison of the algorithms in terms of classification accuracy, classification error, specificity, sensitivity, precision and AUROC revealed that the HLRM was found to be generally the best model for classifying the learners according to their performance in a numeracy test followed by SLRM, CART and Naïve Bayes respectively.

Since the HRLM was found to be the best performing classification algorithm from the current study, the results support its implementation in many previous studies on early childhood numeracy developments such as the ones conducted by [3], [5] and . In addition, the SLRM which has also been used by some previous studies such as the one conducted by [3, 5, 8] and is also useful in classifying learners according to their performance in a numeracy test since it was ranked second after the HLRM. The SLRM also has an advantage over HLRM which is that it does not use all predictor variables since it selects some variables step wisely, but it still has a classification ability which is closer to that of the HLRM. As such, the SLRM may be the algorithm of choice if the researcher's objective is to achieve parsimony.

The HLRM results showed that the relationship between the predictor variables and the learner's performance on a numeracy test is clustered within the interaction between the learner's home language and province. Furthermore, some demographic variables (gender), the family structure variables (staying with father, with grandfather, aunts or with other children) and at home learning activities and family support (getting help from sister) were found to be decreasing the odds of a learner passing the numeracy test. It is worth noting that these variables were examined individually, and it was possible that a learner might have been possessing more than one or all of these attributes at a time and the effect of a combination of these attributes on the learner's numeracy test results was not tested in the current study.

The results of the current study also showed that some family structure variables (staying with mother), home learning activities and family support variables (reading alone more than two times in a week, doing homework more than two times in a week, and getting help with homework more than two times in a week) and basic household infrastructure and resources (having a car, computer, fridge or a washing machine in the household) increase the odds of a learner to pass the numeracy test. However, it should be noted that several other potential predictor variables that were tested in this study were not significant in the HLRM, therefore; they were deemed as not having any significant effect on the learner's performance on a numeracy test.

Based on the results of the current study, it is recommended that future studies should consider adding more classification algorithms to the ones compared in this study, especially the algorithms that have been found to be good classifiers from newly conducted studies but have not been test in a numeracy competence study. In the current study, prior to partitioning the data into training and validation datasets, 71.9% of the whole dataset comprised learners who had failed the test whereas 28.1% were those who passed the test. As such, the difference between the numbers of observations in the event group (passed) and the non-event group (failed) was 43.8%. This is a relatively big difference between the groups of the dependent variable. As such, it is recommended that further research be conducted by varying the ratio of the number of observations in the event and non-event groups of the dependent variable and then comparing the performance of HLRM, SLRM, CART and NB. This will assist future research to determine whether the difference in the number of observations per group of a binary dependent variable has any impact on the performance of these classification algorithms. Another contribution could be to compare the models using imbalanced and balanced datasets (data augmentation methods may come in handy here).

It is also worth noting that the NB have different variants for instance the Tree-augmented NB and the Parent-child Bayesian network [18] whereas only the NB classifier was used in the current study. Also, CART have different ways of being grown such as using the chi-square, CHAID or GINI whereas in the study only Entropy was used [26]. CART also have different ways of being pruned such as the cost complexity [26], but in the current study only the reduce error pruning approach was used. Some other alternatives that were not covered in the current study are the use of backward and forward selection approaches in SLRM as opposed to using the bi-directional selection approach as in the current study. Due the availability of these alternative methods, it is recommending that further studies may be conducted on the classification algorithms that were compared in the current study while taking these alternatives into consideration.

Since the HLRM has been identified as the best algorithm for classifying learners according to their performance on a numeracy test, the HLRM algorithm can be stored may be used to develop a machine that can be used to automatically predict the chances of a child making it in a numeracy test. Knowing the likelihood of the child passing the test given some measured variables can assist the educators and planning personnel to identify at risk learners earlier in the course. These results may be used to inform planning and intervention. When the interventions are administered at an early stage of the leaner's numeracy skills lessons, they may assist the learners who need help to pass the numeracy test.

Based on the significant predictors of the learner's performance in a numeracy test as identified in the current study, it is recommended that where possible the mothers of the learner should stay in the same household as the leaner, the learner should be encouraged to reading alone more than two times in a week and to do homework more than two times in a week, but we recommend the importance of ensuring that the learner gets help with homework more than two times in a week. It is also worth noting that as literature has shown, it is important to have a learner living in a household that have enough resources such as the ones identified in this study which are a car, computer, fridge or a washing machine. All these variables were found to have a positive impact on the learner's chances of passing a numeracy test which is a common measure of childhood numeracy development.

## Acknowledgments

The authors of this research acknowledge the North West University (NWU) for availing its resources to support this research.

## References

[1] R.o.S.A. Department of Basic Education. Snap Survey of Ordinary Schools 1997-2016 [dataset]. Version 1.6. Pretoria: DBE [Producer], Snap Survey of Ordinary Schools 1997-2016 [dataset]. Cape Town: DataFirst [Distributor], (2017).
[2] Y. Anders, H.-G. Rossbach, S. Weinert, S. Ebert, S. Kuger, S. Lehrl, and J. Von Maurice, Home and preschool learning environments and their relations to the development of early numeracy skills, *Early Childhood Research Quarterly,* **27**, 231-244 (2012).
[3] I.K. Lukie, S.-L. Skwarchuk, J.-A. LeFevre, and C. Sowinski, The role of child interests and collaborative parent–child interactions in fostering numeracy and literacy development in Canadian homes, *Early Childhood Education Journal,* **42**, 251-259 (2014).
[4] M.F. Del Río, M.I. Susperreguy, K. Strasser, and V. Salinas, Distinct influences of mothers and fathers on kindergartners' numeracy performance: The role of math anxiety, home numeracy practices, and numeracy expectations, *Early Education and Development,* **28**, 939-955 (2017).
[5] E.A. Frongillo, S. Kulkarni, S. Basnet, and F. de Castro, Family care behaviors and early childhood development in low- and middle-income countries, *Journal of Child and Family Studies,* **26**, 3036-3044 (2017).
[6] D.J. Purpura, E. Day, A.R. Napoli, and S.A. Hart, Identifying domain-general and domain-specific predictors of low mathematics performance: A classification and regression tree analysis, *Journal of Numerical Cognition,* **3**, 365-399 (2017).
[7] S.K. Cheung, X. Yang, K.M. Dulay, and C. McBride, Family and individual variables associated with young Filipino children's numeracy interest and competence, *British Journal of Developmental Psychology,* **36**, 334-353 (2018).
[8] F. Niklas and C. Tayler, Room quality and composition matters: Children's verbal and numeracy abilities in Australian early childhood settings, *Learning and Instruction,* **54**, 114-124 (2018).
[9] S. Phillipson, S.N. Phillipson, and S. Kewalramani, Cultural variability in the educational and learning capitals of Australian families and its relationship with children's numeracy outcomes, *Journal for the Education of the Gifted,* **41**, 348-368 (2018).
[10] P. Aunio, J. Korhonen, L. Ragpot, M. Törmänen, R. Mononen, and E. Henning, Multi-factorial approach to early numeracy—The effects of cognitive skills, language factors and kindergarten attendance on early numeracy performance of South African first graders, *International Journal of Educational Research,* **97**, 65-76 (2019).
[11] L. Mencarini, S. Pasqua, and A. Romiti, Single-mother families and the gender gap in children's time investment and non-cognitive skills, *Review of Economics of the Household,* **17**, 149-176 (2019).
[12] S.A. Reynolds, L.C. Fernald, J. Deardorff, and J.R. Behrman, Family structure and child development in Chile: A longitudinal analysis of household transitions involving fathers and grandparents, *Demographic research,* **38**, 1777 (2018).
[13] T.Y. Hill and N. Palacios, Older sibling contribution to younger children's working memory and cognitive flexibility, *Social Development,* **29**, 57-72 (2020).
[14] T. Kleemans, M. Peeters, E. Segers, and L. Verhoeven, Child and home predictors of early numeracy skills in kindergarten, *Early Childhood Research Quarterly,* **27**, 471-477 (2012).
[15] E. Melhuish, M.B. Phan, K. Sylva, P. Sammons, I. Siraj-Blatchford, and B. Taggart, Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school, *Journal of Social Issues,* **64**, 95-114 (2008).
[16] N.L. Pace, Independent predictors from stepwise logistic regression may be nothing more than publishable P values, *Anesthesia & Analgesia,* **107**, 1775-1778 (2008).
[17] B. Lund, M.A. Solutions, and D. Wilmington, Logistic Model Selection with SAS® PROC's LOGISTIC, HPLOGISTIC, HPGENSELECT, (2017).
[18] Y. Liu, W. Shi, and W. Czika. *Building Bayesian Network Classifiers Using the HPBNET Procedure*, Proceedings of SAS Global Forum, (2017).
[19] C.M.A. Gomes, G.C. Lemos, and E.G. Jelihovschi, Comparing the predictive power of the CART and CTREE

algorithms, *Avaliação Psicológica,* **19**, 87-96 (2020).

[20] E.M. Sledjeski, L.C. Dierker, R. Brigham, and E. Breslin, The use of risk assessment to predict recurrent maltreatment: A classification and regression tree analysis (CART), *Prevention science,* **9**, 28-37 (2008).

[21] Y.-Y. Song and L. Ying, Decision tree methods: applications for classification and prediction, *Shanghai archives of psychiatry,* **27**, 130 (2015).

[22] S. Singh and P. Gupta, Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey, *International Journal of Advanced Information Science and Technology (IJAIST),* **27**, 97-103 (2014).

[23] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, Analysis of various decision tree algorithms for classification in data mining, *International Journal of Computer Applications,* **163**, 15-19 (2017).

[24] H. Kuswanto and R. Mubarok, Classification of Cancer Drug Compounds for Radiation Protection Optimization Using CART, *Procedia Computer Science,* **161**, 458-465 (2019).

[25] N. Patel and S. Upadhyay, Study of various decision tree pruning methods with their empirical comparison in WEKA, *International journal of computer applications,* **60**, (2012).

[26] SAS Institute Inc. *SAS/STAT® 14.1 User's Guide*, Cary, NC, (2015).

[27] Q. Wei. *Understanding of the naive Bayes classifier in spam filtering*, AIP Conference Proceedings, 020007, (2018).

[28] S. Misra, H. Li, and J. He. *Noninvasive fracture characterization based on the classification of sonic wave travel times*, in Machine Learning for Subsurface Characterization, Gulf Professional Publishing, 243-287, (2020).

[29] S. Krishnan, *Biomedical Signal Analysis for Connected Healthcare*, Elsevier(2021)

[30] J.W. Osborne, Advantages of hierarchical linear modeling, *Practical Assessment, Research, and Evaluation,* **7**, 1 (2000).

[31] Y. Tang, Y.-Q. Zhang, N.V. Chawla, and S. Krasser, SVMs modeling for highly imbalanced classification, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* **39**, 281-288 (2008).

[32] A. Tharwat, Classification assessment methods. Applied Computing and Informatics, 17 (1), 168–192. 2020.

[33] J. Huang and C.X. Ling, Using AUC and accuracy in evaluating learning algorithms, *IEEE Transactions on knowledge and Data Engineering,* **17**, 299-310 (2005).