A Predictive Model for Standardized Test Performance in Michigan Schools

William Sullivan, Joseph Marr and Gongzhu Hu

Abstract Public school officials are charged with ensuring that students receive a strong fundamental education. One tool used to test school efficacy is the standardized test. In this paper, we build a predictive model as an early warning system for schools that may fall below the state average in building level average proficiency in the Michigan Educational Assessment Program (MEAP). We utilize data mining techniques to develop various decision tree models and logistic regression models, and found that the decision tree model with entropy impurity measure accurately predicts school performance.

Keywords Predictive modeling · Decision tree · Logistic regression · MEAP proficiency

1 Introduction

The ability of governing bodies to hold schools and school districts accountable to standardized test scores has been the subject of heated debate for many decades. Michigan adopted the Michigan Educational Assessment Program (MEAP) at the beginning of the 1969–1970 school year and administered the MEAP until 2013–2014. It was replaced with the Michigan Student Test of Educational Progress in 2015.

The MEAP test was taken every fall by public school students in grades 3–9. The test measured student proficiency in a number of different subjects. Due to budgetary constraints the only two subjects, reading and mathematics, in which every student

W. Sullivan · J. Marr · G. Hu (✉)
Department of Computer Science, Central Michigan University,
Mount Pleasant 48859, USA
e-mail: hu1g@cmich.edu

W. Sullivan
e-mail: sulli2wf@cmich.edu

J. Marr
e-mail: marr2ja@gmail.com

© Springer International Publishing AG 2017

R. Lee (ed.), *Applied Computing and Information Technology*,
Studies in Computational Intelligence 695,
DOI 10.1007/978-3-319-51472-7_3

is tested every school year. The MEAP test scores proficiency on a four point scale in which a score of 1 is highest and 4 is lowest. Students who score 1 or 2 are considered proficient and students who score 3 or 4 are considered not proficient.

Michigan holds schools accountable to the building level test scores of their students. The state tracks each schools percentage proficient and holds schools that fall below the state average accountable by either assigning them a “priority” ranking or by requiring “action plans” to help improve their test scores. In recent years state funding has been tied to student test scores as a way to increase levels of proficiency.

The aim of this research is to develop a predictive model that calculates building level proficiency scores of reading and mathematics and then predicts using different demographic factors the likelihood a school will achieve a proficiency percentage above the state average.

2 Related Work

The application of data mining to educational research is a relatively new endeavor. The field of Educational Data Mining (EDM), in general, is characterized by traditional data mining techniques and the inclusion of psychometric methodology [1, 2]. Much of the EDM literature has focused on improving student learning models or studying pedagogical support of learning software [3–5].

However, the application of classification techniques have also been applied to education research in the form of so called “Early Warning Systems”. These systems have primarily been concerned with predicting high school drop outs. Bowers, Sprott, and Taff reviewed 110 proposed models and suggest standardized metrics for evaluation using latent class models [6]. Carl et al. developed a regression based “Early Warning System” which applies to a broader set of outcomes beyond dropping out of school [7]. Baradwaj and Pal focused on predicting student performance in higher education [8]. Knowles et al. take things a step further and utilize a number of ensemble methods including boosted tree and neural network frameworks [9].

In our study we built logit regression models and decision tree models to predict the students’ performance in the MEAP tests.

3 Data

The data for this research comes from a number of different public sources:

(a) Michigan Student Data System (MSDS) [10]. The data include school types (traditional or charter), delivery method (traditional or virtual), etc.
(b) Michigan School Data [11]. The data set include teach/student ratio, Free and Reduced Lunch eligibility, and MEAP Proficiency scores by grade for all public schools in Michigan.

Table 1 Descriptive statistics of variables

Variable	Statistic				
	N	Mean	St. dev.	Min	Max
No. proficient	1,541	306	266	1	1,855
Percent proficient	1,541	0.510	0.185	0.023	0.989
Crime rate (CR)	1,541	0.065	0.053	0	1.117
Violent CR	1,541	0.006	0.006	0	0.025
Share of violent crime	1,541	0.079	0.050	0	0.269
% of population HS education or higher	1,541	87.251	6.041	67.300	99.400
Income	1,541	32,123	10,197	12,295	101,402
Free and reduced lunch (% eligible)	1,541	0.565	0.265	0.017	1.000
Free-lunch (% eligible)	1,541	0.510	0.267	0.012	1.000
Share of free-lunch	1,541	0.879	0.082	0.496	1.000
Teacher/student ratio	1,541	16.065	5.023	4.500	169.000
Jobless rate	1,541	7.520	1.818	4.700	12.700
State average (% proficient)	1,541	0.545	NA	NA	NA
Charter	218	NA	NA	0	1
Big 4	301	NA	NA	0	1

(c) Uniform Crime Report [12]. The report contains the total number of crimes along with details in over 350 cities in Michigan, including crime rate and share of violent crimes.

(d) American Community Survey (ACS) [13]. This survey provides information about jobless rate, education and income data for cities in Michigan.

From these data sets, we merged building level enrollment, demographic and income data with city level educational attainment and crime rate data to produce a sample of over 1,500 observations. This data set was used to predict the likelihood that a school achieves a proficiency percentage greater than the state average. The descriptive statistics of the variables are given in Table 1. In the following we shall describe these variables in more details.

3.1 School Proficiency Percentage

The Michigan Educational Assessment Program (MEAP) test was administered every fall to students in grades 3 through 9 from 1970–2014. The test evaluates a student's knowledge of the material covered the year before. The student is scored on a four-point scale where a score of 1 or 2 is deemed "proficient." The state uses

Table 2 Proficiency comparison of sample and population

Summary statistics	N	Mean
State percent proficiency	3,159	0.545
Sample percent proficiency	1,541	0.511

the percentage of students proficient as one of many ways to measure whether or not a school is effective. The percentage of proficiency of the population and sample is given in Table 2.

3.2 Crime Rate

The annual Uniform Crime Report (UCR) published by FBI provides detailed population and crime data on more than 350 cities in Michigan for 2014. For this research, it is apparent that when the data set is trimmed to include only schools located within cities in which UCR data is available the proficiency percentages of the sample are representative of the population as a whole. This results in a sample of just 1,541 observations.

Three crime rates (CR) are calculated:

$$CR = N_{crime}/N_{population}$$
$$Violent\ CR = N_{violent\ crime}/N_{population}$$
$$Share\ of\ Violent\ CR = N_{violent\ crime}/N_{crime}$$

where N_x is the number of x. These measures help separate the likely correlation between crime rate and other independent variables.

3.3 Education

The American Community Survey (ACS) is sent out to 3.5 million households each year and with this data the United States Census Bureau creates estimates of income and education. The estimates that were selected for the purposes of this research are estimates of the percentage of the population within a city that is 25 years or older and has obtained a high-school diploma or higher. The data from the ACS is then matched to each city with UCR data available resulting in a data set with 1,541 observations.

Table 3 Free and reduced lunch eligibility thresholds

FRL thresholds	Max income eligible (family of 4)	% in relation to poverty line
Free	$30,615	135
Reduced	$43,568	185

3.4 Income

The Michigan Department of Education (MDE) publishes a report detailing the number of students eligible for free and reduced lunch for each school every year. Table 3 shows the free and reduced lunch eligibility for 2014.

The percentage of students eligible for free and reduced lunch is used as a proxy for city level income data. This is appropriate because the student's eligibility is directly related to their family's income.

3.5 Class Size

Class size as measured by the student-to-teacher ratio for a school was calculated using the MDE's annual report on educator effectiveness and student count. It is traditionally considered an indicator of school quality. A number of schools are excluded because their primary method of delivery is online or they offer a significant number of online courses. Including these outliers would skew the data.

3.6 Jobless Rate

The number of people without a job has long been an economic indicator of great importance [14]. The ACS provides estimates for the jobless rate in every city in Michigan. The jobless rate of the city has been merged with all of the other demographic and building level information to help control for as many demographic variables as can be observed.

3.7 Charter Schools & Big 4 School Districts

A few dummy variables were created regarding the types of schools and their locations. One variable indicates whether or not a school is located in one of the "big 4" districts (Detroit, Lansing, Grand Rapids, Flint) in Michigan. Policy makers have long used these districts as a baseline in comparison to the others. The results and

Table 4 Sample comparison to charter schools and big-4

Summary statistics	N	Mean proficiency %
Charter public schools	218	37.2
Traditional public schools	1,249	53.4
Big 4 (traditional and charter)	213	29.7
Big 4 (charter only)	97	29.2

atmosphere of schools within these larger districts are considered more complex than the rest of the state. This may provide some interesting insight into the difference between schools that operate in an urban Michigan setting and schools that operate in a rural area. The other dummy variable was to identify charter schools. Table 4 shows the comparison of charter schools and "big 4."

3.8 *Target Variable*

State agencies often classify schools as underperforming or performing by whether or not performance is above or below the state average. The state average for school proficiency in our data is 54.5%. Thus, we generate a target variable called "Above Average" which takes on the value 1 for schools whose percent proficient is equal to or greater than 54.5% and 0 otherwise. Figure 1 provides a distribution of percentage proficient by schools. The distribution is mostly normal. Scores to the right of the mean (red line) are classified above average.

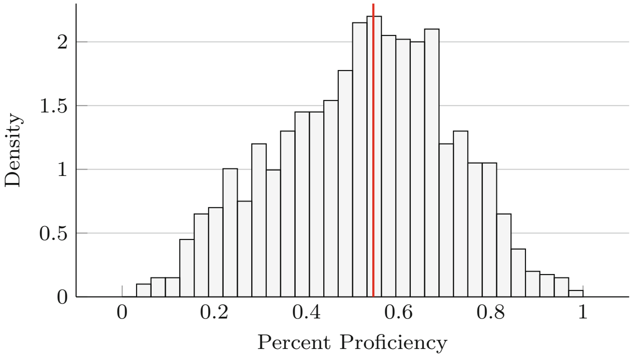
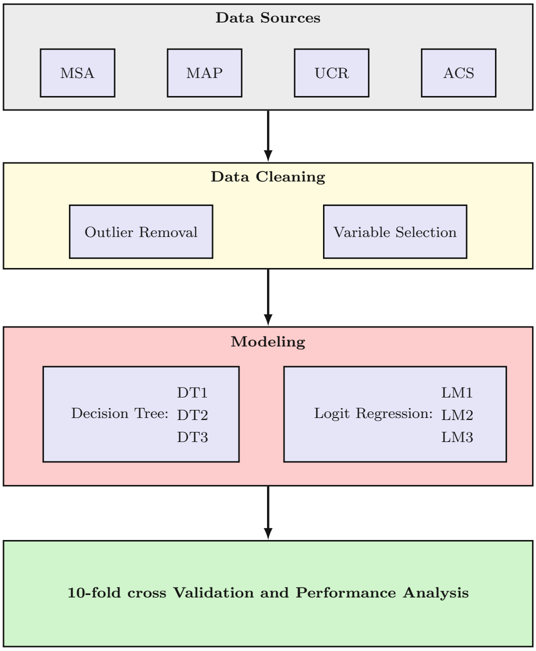


Fig. 1 Distribution—percentage proficient

Fig. 2 Work flow


Data Sources
MSA
MAP
UCR
ACS
Data Cleaning
Outlier Removal
Variable Selection
Modeling
Decision Tree: DT1 DT2 DT3
Logit Regression: LM1 LM2 LM3
10-fold cross Validation and Performance Analysis

4 Methodology

The overall work flow of our analysis is shown in Fig. 2. We have described the data sources in the previous section, and shall discuss each of the remaining steps in this section.

4.1 Data Cleaning

Before analyzing the data, the data need to be trimmed and cleaned to include only schools operating within a city where all demographic information was available. The summary statistics show that the statewide proficiency percentage was within 3% of the samples average proficiency percentage. This means that the sample is representative of the population.

Prior to modeling, a filtering was applied to the data to remove observations with extreme outliers. As an example, one observation reported a Student/Teacher Ratio of 168, in contrast to the normal ratio approximately 17 for traditional classrooms. We believe this observation represents a virtual classroom. In total, 51 observations were filtered from the dataset with $n = 1490$.

Variable filtering was also applied. Any variable directly used to calculate our target variable was excluded from further consideration. This step eliminated variables *Number Proficient*, *Percent Proficient*, and *State Average* (*% Proficient*) with 12 input variables remaining.

4.2 Modeling

We employed SAS Enterprise Miner 13.2 to build and analyze our models. Two primary model types were considered, Decision Tree and Logit Regression. Each model type was tested with a variety of splitting rules, variable selection criteria, and variable transformations.

4.3 Decision Trees

A decision tree is a rule-based modeling technique [15]. The basic idea is to split the data set (node in the tree) into subsets (children nodes) based on a splitting criterion on the relationship of the target variable with the input variables. The splitting process continues for each child node until certain condition is met.

Three decision rule models are developed for our study, as listed below.

DT1: Decision Tree with Chi-Square splitting rule, no variable transformation
DT2: Decision Tree with Entropy splitting rule, no variable transformation
DT3: Decision Tree with Chi-Square splitting rule, transformed variables

4.3.1 Chi-Square Splitting

The Chi-square splitting process requires considering a number of split points for each input variable. It is performed on each binary split. The optimal split is selected via the *logworth* value. Bonferroni adjustment is applied to compare *logworth* to $-\log(\alpha/m)$ for a predetermined significance level α and number of comparisons m. The Chi-square splitting process is given in Algorithm 1.

We also applied tree pruning to optimize predictive power and parsimony, and specified max tree depth and min size of tree nodes to improve the run-time performance.

4.3.2 Entropy Splitting

Our second decision tree model (DT2) is based on Entropy splitting rules. Entropy for a given node j is defined as

Algorithm 1: Chi-Square Splitting *SplitAttribute*(*t*)

```
Input: Tree node $t$ with variables $v_i$'s
Input: $\alpha$ significance level
Input: $m$ number of comparisons
Output: $j$ – the splitting variable
begin
    foreach $v_i$ do
        $p_i \leftarrow \chi_i^2(t)$
        $logworth_i \leftarrow -\log(p_i)$
        if $logworth_i > -\log(\alpha/m)$ then
            $v_i$ is marked significant
        end
    end
    $logworth_j \leftarrow \max_i(logworth_i)$ for $v_i$ marked
    return $j$
end
```

$$H(j) = -\sum_{i=1}^{k} P_i \log_2 P_i$$

where P_i is the posterior probability of the child node i of node j. Essentially, we are interested in pure child nodes or nodes which are as close as we can get to either Above Average = 1 or Below Average = 0. The node being considered for split contains both values and we want the split that maximizes child node purity. When considering our possible splits we calculate the reduction in our impurity index (in this case, Entropy). This calculation takes the form

$$H(j) - \sum_{i}^{r} P_i * H(i)$$

where r is the number of child nodes of j. This process continues until a minimum reduction in impurity is met. We used the default settings in SAS Enterprise Miner for this threshold. Pruning of the candidate tree follows the same procedure as outlined above for the Chi-square selection. Maximum tree depth was set to 6 (7 total levels).

4.3.3 Variable Transformation

Our third decision tree (DT3) includes a variable transformation node prior to the decision tree node. The variables are transformed according to variable type, binary and interval in our data. Binary variables do not require any transformations. Interval variables may, however, fit our modeling process better if they are transformed. The "Best" option in Enterprise Miner was selected that attempts multiple methods. The two primary methods are optimal binning and a power transformation which mimics the so-called Box-Cox power ladder to maximize normality.

4.4 Logit Regression

Logit (or Logistic) regression models [16] are a type of general linear regression with binary target variable. Our target variable is binary. Therefore, a nonlinear regression model is called for. The logit regression model differs from general linear equation in the following way. A general linear regression can be expressed as

$$y = \beta_0 + \sum(\beta_i x_i) + \epsilon$$

where y is a continuous dependent variable and x_i represents independent variables which may be either continuous or binary. In our case the dependent variable is not continuous and the probability of $y = 1$ is estimated:

$$P(y = 1) = \frac{1}{1 + \exp(-(\beta_0 + \sum(\beta_i x_i)}$$

The regression coefficients β_i can be exponentiated to determine the odds of y for a change in x_i. The three logit regression models are listed below.

LM1: Logit model with no variable transformations
LM2: Logit model with variable selection performed by decision tree (Chi-Square)
LM3: Logit model with variables transformed and then selected based on Chi-Square and R-Square

The process of selecting variables for an optimal logit model can be conducted in a number of ways. We employ stepwise selection for all three logit models. The general process is outlined in Algorithm 2.

Algorithm 2: Variable Selection for Logit Reg Models

```
Input: Data set $D$ with variables $v_i$'s
Input: $\alpha$ significance level (e.g. 0.05)
Output: $V$ – set of variables selected for the model
begin
    $V \leftarrow \{v_j\}$ where $v_j$ has the max$(R^2)$
    foreach $v_i$ do
        $p_i \leftarrow \chi_i^2(D)$
        if $p_i > \alpha$ then
            $V \leftarrow V \cup \{v_i\}$
        end
    end
    return $V$
end
```

4.4.1 Variable Selection

Our second logit model (LM2) employs a decision tree node before the logit model node to select candidate variables. The decision tree is “grown” via Chi-square splitting in the same manner outlined in Sect. 4.3.1. In this case, however, the tree is allowed to grow until the *logworth* value no longer exceeds the threshold. The tree is not pruned but instead the resulting variables are passed to the logit node for consideration in the model.

Variable selection for model 3 (LM3) is conducted post variable transformation. We employed an R^2 selection method. This step consisted of comparing each variable (including transformed versions) in a forward selection process against a minimum R^2 threshold (0.005).

4.4.2 Variable Transformation

In logit model 3 (LM3), we also include a variable transformation step before variable selection. This step follows the same process outlined in Sect. 4.3.3 which was used to transform variables prior to decision tree model DT3.

4.5 Validation

Due to the small number of observations in our dataset, partitioning the data into training, validation, and test sets would be inappropriate. Instead, we employed k-folds cross-validation with k set to 10. The data set is partitioned into k equal sized random samples. One sample is held as a test set while the others are used as training set to build the models. The process of training and testing is repeated until all k folds have an opportunity to serve as the test set. The results are then averaged and taken as final.

5 Experimental Results

5.1 Final Model Selection

Our model selection process mimicked a bracketed single elimination tournament. Two brackets were formed which consisted of the decision tree models on one side and the logit models on the other. The decision tree models were selected based on the lowest misclassification rate while the logit models were selected based on the lowest average squared error. Once the winning model from each bracket was selected the overall winner was chosen based on lowest average squared error.

Table 5 All decision trees—fit statistics

Model #	Misclassification rate	Ave. squared error
DT2*	0.140268	0.101133
DT3	0.151678	0.108770
DT1	0.159732	0.133661

∗Model selected

Table 6 Logit regression—fit statistics

Model #	Misclassification rate
LM3*	0.107184
LM2	0.109252
LM1	0.109252

∗Model selected

The best decision tree model was the Entropy model (DT2). Table 5 summarizes the misclassification rates and the average squared error for all three decision tree models on the target variable *above average proficiency*. Misclassification rates are within 2% points of one another. The range of the average squared error was 0.032528.

Among the three logit regression models, the best is LM3, the model with variable transformation and variable selection employed. This outcome is not surprising. We might expect the model with variables optimized via the transformation process to be the best performer. However, the range of the average squared errors for the three models is 0.002068. This is a much smaller range than we found in our decision tree models that is 0.0325. Table 6 summarizes the misclassification rates for each model.

All three logit models perform relatively similar. Transformation had little effect on performance.

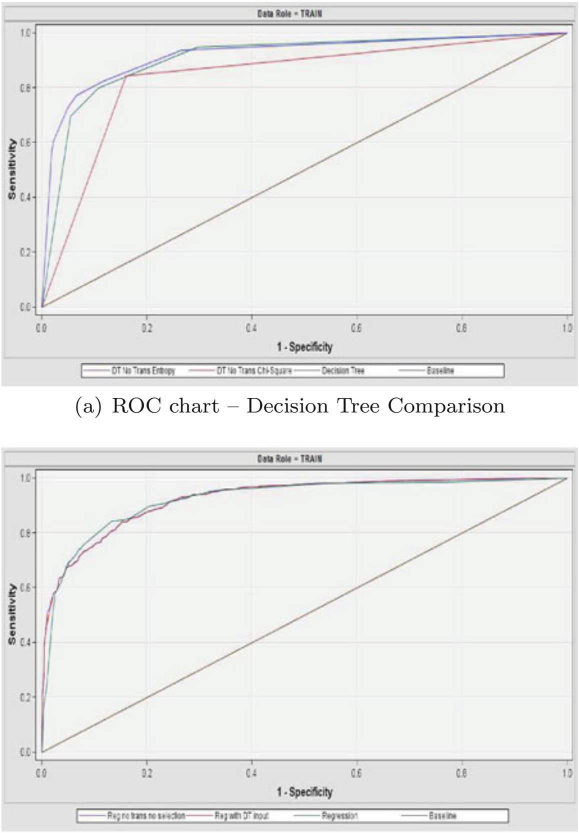
The Receiver Operator Curve (ROC) chart of the decision tree models and for the logit regression models are shown in Fig. 3a, b respectively.

The final model selected was the entropy based decision tree (DT2). Based on average squared error, this model performed slightly better than the best logit model (LM3). The average squared error for the decision tree model was 0.101133. The average squared error for the logit model was 0.107184. The difference, 0.006051, is small even when compared to the tight range of average squared error for all three logit models. Nonetheless, the decision tree's performance is better.

The full entropy decision tree model is shown in Fig. 4. It begins by splitting on the percentage of students receiving free lunch. Approximately 46% of the sample is classified at a terminal child node (node 3). This implies that the free lunch variable is a strong predictor of school proficiency.

The next split occurs on percentage of students receiving free or reduced lunch. Again, we see a strong prediction of school proficiency. Of the remaining schools and those with less than 39% of students receiving free or reduced lunch, 96% scored above average MEAP proficiency.

Fig. 3 ROC charts of decision trees and logit regression



(a) ROC chart – Decision Tree Comparison

(b) ROC chart – Logit Comparison

The remainder of the 454 schools are split again by free and reduced lunch with the poorest districts being further defined by jobless rate, share of violent crime and overall violent crime rate. These results indicate that our model conforms to known theoretical and empirical research on school performance. The simplicity of the model also provides ease of explanation to policy makers and stake holders.

5.2 Model Performance

Table 7 reports fit statistics for the winning model (DT2). The ROC index of 0.92 is well over the common threshold of 0.80 as shown in Fig. 3a. This compares favorably with the early warning systems reviewed by Bowers et al. The majority of those models had values below 0.75. In fairness, many of those models were attempting to predict more complicated outcomes.

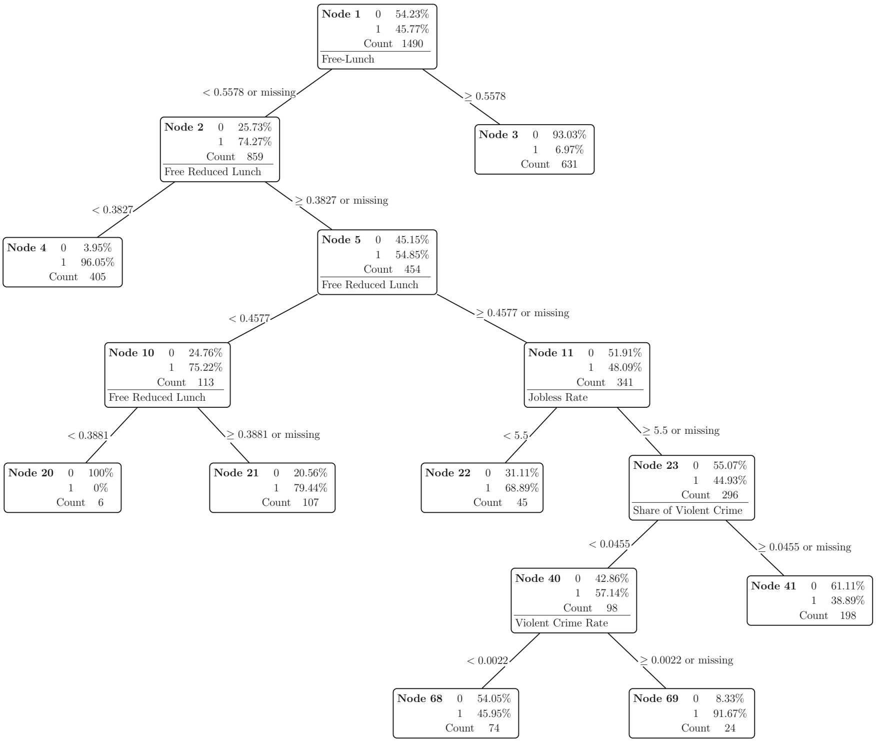

Node 1 0 54.23% 1 45.77% Count 1490
Free-Lunch
< 0.5578 or missing
≥ 0.5578
Node 2 0 25.73% 1 74.27% Count 859
Free Reduced Lunch
Node 3 0 93.03% 1 6.97% Count 631
< 0.3827
≥ 0.3827 or missing
Node 4 0 3.95% 1 96.05% Count 405
Node 5 0 45.15% 1 54.85% Count 454
Free Reduced Lunch
< 0.4577
≥ 0.4577 or missing
Node 10 0 24.76% 1 75.22% Count 113
Free Reduced Lunch
Node 11 0 51.91% 1 48.09% Count 341
Jobless Rate
< 0.3881
≥ 0.3881 or missing
< 5.5
≥ 5.5 or missing
Node 20 0 100% 1 0% Count 6
Node 21 0 20.56% 1 79.44% Count 107
Node 22 0 31.11% 1 68.89% Count 45
Node 23 0 55.07% 1 44.93% Count 296
Share of Violent Crime
< 0.0455
≥ 0.0455 or missing
Node 40 0 42.86% 1 57.14% Count 98
Violent Crime Rate
Node 41 0 61.11% 1 38.89% Count 198
< 0.0022
≥ 0.0022 or missing
Node 68 0 54.05% 1 45.95% Count 74
Node 69 0 8.33% 1 91.67% Count 24

Fig. 4 The entropy decision tree model

Table 7 DT2—fit statistics

Fit statistics	Value
Misclassification rate	0.14
ROC index	0.92
Average squared error	0.10
Cumulative lift	2.10
Gain	109.84

The misclassification rate (0.14) is reasonable but further inspection via a confusion matrix (Table 8) and a misclassification chart (Fig. 5) indicates that our classification of Above Average is less accurate than for schools below the mark. In other words, we have a larger proportion of false positives than we do false negatives. This is a concern. If a school is classified as performing above the average but in reality they are below they may not receive the proper policy prescription.

The goal of the model is predict which schools will fall below the average. Our accuracy in this regard is strong. Less than 10% of schools who are classified as below average are actually above. Again, they are the schools that fall into the false positive category that should be cause for concern.

Table 8 Confusion matrix—DT2

n = 1490		Predicted	
		Below average	Above average
Actual	Below average	754	155
	Above average	54	527

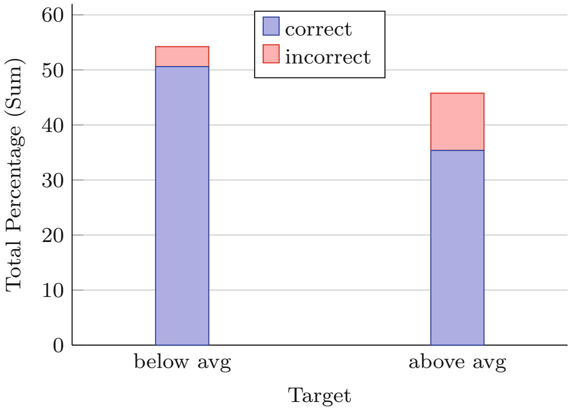


Fig. 5 Misclassification chart—DT2

A number of steps may help improve the accuracy of our model. First, the model may be sensitive to the cutoff point. A large number of schools are centered around the mean average proficiency. While this cutoff may be optimal from the state's perspective, it may be not be the optimal point for classification.

We did not employ any interaction terms in any of our models. It may be beneficial to allow, for instance, Big 4 (urban schools) to interact with Charter. Or, perhaps, crime rate interacts with charter schools differently than public schools.

6 Conclusion

Our final model, a decision tree with entropy splitting, does an adequate job of classifying schools as either above or below the state proficiency average. The model compares favorably with other early warning systems in the EDM literature. The tree utilizes a number of socioeconomic variables including percentage of students receiving free lunch, percentage of students receiving free or reduced lunch, jobless rate, violent crime rate, and share of violent crime. The model does not employ any school quality variables such as student-to-teacher ratio, or class size. Class size is traditionally considered a significant factor indicative of student performance, however in recent years research suggests that the characteristics of the students in the classroom is far more important than the number of students in the classroom [17].

Though there is conflicting research regarding this metric, our analysis showed that it is not as important as other Socioeconomic factors.

Additional model types should also be considered in future research. Individual student data is often included in school performance models. The typical approach is a multilevel model approach, one model for the student level and one for the school level. The inclusion of student level data, along with the requisite modeling techniques, may provide further improvement on classification performance.

References

1. Ryan S.J.d. Baker. Data mining for education. *International encyclopedia of education*, 7:112–118, 2010.
2. Alejandro Peña-Ayala. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4):1432–1462, 2014.
3. Ryan Shaun Baker, Albert T Corbett, and Kenneth R Koedinger. Detecting student misuse of intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*, pages 531–540. Springer, 2004.
4. Joseph E Beck. Engagement tracing: using response times to model student disengagement. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, pages 88–95, 2005.
5. R Charles Murray and Kurt VanLehn. Effects of dissuading unnecessary help requests while providing proactive help. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, pages 887–889, 2005.
6. Alex J Bowers, Ryan Sprott, and Sherry A Taff. Do we know who will drop out?: A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal*, 96(2):77–100, 2013.
7. Bradley Carl, Jed T Richardson, Emily Cheng, HeeJin Kim, and Robert H Meyer. Theory and application of early warning systems for high school and beyond. *Journal of Education for Students Placed at Risk (JESPAR)*, 18(1):29–49, 2013.
8. Brijesh Kumar Baradwaj and Saurabh Pal. Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2(6):63–69, 2011.
9. Jared E Knowles. Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining*, 7(3):18–67, 2015.
10. Center for Educational Performance and Information. Michigan School Data. http://www.michigan.gov/cepi.
11. Michigan Department of Education. Michigan School Data. https://www.mischooldata.org.
12. Federal Bureau of Investigation. Uniform Crime Report. https://www.fbi.gov/about-us/cjis/ucr/ucr.
13. United States Census Bureau. American Community Survey (ACS). https://www.census.gov/programs-surveys/acs/data.html.
14. Theodore G Chiricos. Rates of crime and unemployment: An analysis of aggregate research evidence. *Social problems*, 34(2):187–212, 1987.
15. J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.
16. Frank Harrell. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.
17. Caroline M Hoxby. The effects of class size and composition on student achievement: new evidence from natural population variation. Technical report, National Bureau of Economic Research, 1998.