Integration of Artificial Intelligence and Rough Set Methodology
to Engineer a Predictive School Classification System

Nina Kajji, Gordon H. Dash, Elliot Krieger
Integration of Artificial Intelligence and Rough Set Methodology to Engineer a Predictive School Classification System

Nina Kajiji\textsuperscript{1}, Gordon H. Dash, Jr.\textsuperscript{2}, Elliot Krieger\textsuperscript{3}

\textsuperscript{1}The NKD-Group, Inc. and Department of Computer Science and Statistics, University of Rhode Island
\textsuperscript{2}Finance and Decision Sciences Area, College of Business Administration, University of Rhode Island
\textsuperscript{3}Office of the Commissioner, Rhode Island Department of Elementary and Secondary Education

Contact: nina@uri.edu

Abstract

This research cross contemplates two artificial neural network algorithms to engineer a predictive assessment-based classification system for the objective reporting of ranked high school achievement in a manner that is consistent with the national (United States) mandated No Child Left Behind Act (NCLB). The research interrogates how effectively five indexes (learning score indicators – LSIs) classify public high schools in a manner that is consistent with NCLB guidelines. It is a widely held belief that public school administrators should incorporate these learning indicators in their development of formative action items that directly impact the assessment of overall school achievement. The study proceeds by describing the high school classification problem within the context of a rough set. Multicriteria decision aiding (MCDA) begins by producing an ANN driven objective classification of Rhode Island high schools based on ex-post year-to-year assessment differentials. This is followed by an empirical investigation that uses a mapping ANN topology to engineer a prediction probability for next period NCLB classification. We find the five LSI measures to be efficient classifiers of ex-post school achievement in AY2004-2005. Moreover, we provide evidence to show that when applied to out-of-sample study data for AY2005-2006, the MCDA probabilistic predictive model and the subjectively derived actual reclassifications aligned with measurable accuracy.
1. Introduction

The global economy has produced rapid changes in both the nature and complexity of what has become a more technologically demanding work place. The relaxation of national borders within this evolving global economy has foisted new and increased demands onto the world’s educational systems. However, from the most advanced economies to those that are still evolving there is a perception that educational production economics has failed to identify regional production functions that achieve the goal of producing maximum output given a defined input combination. This production inefficiency has resulted in delayed development of technological skills as well as diminished individual student preparedness. Of course, it is a widely held view that education, particularly high-school education, is a critical component of preparing citizens for global economic competition. As such, many nations have taken the results of international assessments seriously enough to institute administrative changes that are designed to narrow achievement disparities, both national and global, by seeking to produce high performing schools. While the debate continues on how to best define a high-performing school, the extant literature suggests that, at a minimum, high performing schools focus on prioritizing student achievement, coherent and standards-based curriculum, the use of data, and the availability of basic instructional resources.

In the United States, part the response to this global challenge was the passage of the federal “No Child Left Behind Act of 2001” (NCLB) which is administered by the U.S. Department of Education (ED). Among its many objectives, NCLB requires every state to measure English language arts, mathematics, and science achievement at the elementary-, middle- and high-school levels. Importantly, NCLB interpretations could cause the issuance of directives to those schools that miss annual targets. Such directives might include instructions to use some federal aid for specific school improvement initiatives, such as school choice or supplemental education services. It is the ED’s Strategic Plan for Fiscal Years 2007-12 (Plan) that provides a comprehensive summary of how federal support is targeted for deployment to encourage a move towards excellence and equal educational access on a national level (see, Goals [1]; NCLB [2]). Moreover, the ED also relies on the Plan to assist states with development of uniform accountability standards to guide the interpretation of ex-post performance at the school level. In addition to requiring individual states to objectively classify school performance based on prior-to-current year performance improvement, the ED assessment process considers other key factors. One key factor is a comparison of how closely individual schools / districts met the prior year targets by whole school and individual group constituency (e.g., Black, Hispanic, Asian, Native American and White). When a school or a district falls below the annual target the administrative unit may get credit for meeting the target if it is deemed that sufficient progress has been made towards the goal. In the state of Rhode Island these targets are for reading (English-language arts), mathematics, graduation rates (for high schools), and one other indicator as selected by the state (attendance).

Over time, the operational goals established by the ED Plan are expected to ensure that an increasing number of schools / districts meet their annual goals. That is, these educational units make adequate yearly progress. However, objective and efficient operational command over

---

assessment evaluation is the crucial step where decision-makers still encounter the need for efficient procedures to achieve uniform and comparable rankings in a statewide system.

1.1 *Multicriteria Decision Analysis for Efficient School Assessment*

For school administrators, the task of assigning individual school units to predefined homogeneous performance groups is a complex process. The overall complexity of the process is exacerbated by the use of relative judgments on how to best weight alternative summative assessment measurements for the various school units. Summative assessments are comprised of periodic examinations to determine what students know and do not know. But, because school units are established independently from the alternative measurement instruments under consideration, the classification and subsequent sorting of the observed measurements is best accomplished by a method that yields some comparison to an objectively determined reference profile that is capable of distinguishing among the groups (absolute judgments). Multicriteria decision analytics (MCDA) is a systematic decision-tool that can provide a formal statement for both classification and sorting problems (for a review, see Zopounidis and Doumpos, [3]). Specifically, the process of nominally aggregating school units into groups is representative of the MCDA classification and sorting objectives.

1.2 *MCDA for Efficient Public School Classification*

We argue here that the practical application of classifying schools into performance groups entails both a classification problem (pattern recognition of school attributes) and a sorting problem (obtaining efficient ranked-prediction probability of within group assignment). By way of example, the state of Rhode Island school performance classification system for academic year 2005 (AY2004-2005) was accomplished by an administrative review of various performance data that featured test-measured achievement of the enrolled students’ in mathematics and English-language-arts proficiency. The R.I. Department of Elementary and Secondary Education (RIDE) used this review to classify individual school units into one of three performance categories: *High Performing*, *Moderately Performing*, or *In Need of Improvement*.  

To highlight the intersection of ED guidelines and the Rhode Island approach, consider the ramifications of a school classified *In Need of Improvement*. In the evaluation year, poor performance or inadequate improvement in either English or mathematics (or both) was a recipe for a school to be classified as *In Need of Improvement*. Whenever a school is so labeled for two or more years, state administrators are obligated to corrective actions. To avoid facing corrective action or intervention school administrators have called upon state officials to enumerate a set of prioritized attributes that have the best chance of producing statistically measurable change in the performance of students assigned to the school. Recent findings have reported that a failure by state-wide policymakers to develop such a response can lead to school administrators viewing accountability measures as punitive rather than constructive (For a more detailed review, see Jamentz, [4]).

---

2 Recently RIDE adopted a classification system that identifies schools as making (not making) adequate yearly progress.
The Rhode Island response to this challenge led to a collaboration between RIDE and the National Center on Public Education and Social Policy (NCPE). In 2002, the two entities identified five school learning support indicator indices (LSIs) for uniform use across public schools at the elementary-, middle- and high-school levels. A primary design goal of the LSI index system was to provide administrators with a measure to gauge whether schools were on track toward improvement even before the improvement was confirmed by rising test scores. For state policymakers, the index based approach was found to avoid the common mistake of weighing down the performance / accountability system (See Ananda and Rabinowitz [5] for a discussion of the ramifications of over-specifying the evaluation process).

1.3 Research Objectives and Organization

The objective of this research paper is twofold. The first objective is to efficiently classify Rhode Island high schools by using an artificial neural network to data mine a state-wide database for nontrivial latent information that may prove useful in the efficient classification of schools as statutorily required. The classification experiment encompasses the five endogenously produced LSI scores as published in the State Report Card. The second objective is to develop a predictive probabilistic classification system by engineering a different, but congruent, artificial neural network to predict a sorted placement of schools based on the latent relationships found within the five LSI measures and the state defined NCLB-inspired performance classification system.

The remainder of the paper is organized as follows. Section 2 presents the ANN techniques that are used in the study. This includes the Kohonen self-organizing map (K-SOM) for efficient ex-post classification of Rhode Island high-schools and the K4 Radial Basis Function (RBF) ANN method for probabilistic prediction. Section 3 presents the data and associated transformation related issues. This is followed by a presentation of modeling results in section 4. Section 5 offers a summary and a conclusion.

2. Artificial Neural Networks for Data Mining

Data mining is a critical activity for educational policy units with complex non-linear databases. Data mining may be described as the interactive process of discovery that emphasizes a search for latent patterns (features) that are not easily detected in student, family and school databases that occur over time. Once these patterns are known, the important features can be examined for policy-making implications. That is, it is possible to interrogate the relative contribution of features to the desired outcome. Policy redirection may also be indicated as once the relative contributions are identified it may then be possible to impute what adjustments, if any, are needed to improve the overall performance of the production system. Methodologies for feature extraction and prediction are discussed below.

---

3 In 2009, NCPE accepted a new mission and was renamed The Center for School Improvement and Educational Policy.

4 In 2005, the Health Knowledge and Skills indicator was dropped due to budget considerations.
2.1 Rough Sets, MCDA and Efficient School Classification

The rough sets theory first introduced by Pawlak [7-8] has a proven track record as an excellent mathematical tool for the analysis of objects that are described by inconsistent, ambiguous or otherwise inferior information. The importance of this work lies in the fact that rough sets assume that every object has some amount of information attached to it. Objects with the same description are indiscernible with respect to the available information. However, a partition is reached when indiscernible objects are cast into their own groups. In a general view of rough sets, Greco, et. al. [9] describes the basic organization of information regarding the objects under analysis. Such data is presented in a data table as the 4-tuple set \( S = \langle U, Q, V, f \rangle \), where \( U \) is a finite set of objects (school units), \( Q = \{q_1, q_2, ..., q_m\} \) is a finite set of attributes (in this case LSI measures), \( V_q \) is the domain of the attribute \( a \), \( V = \bigcup_{q \in Q} V_q \) and \( f: U \times Q \rightarrow V \) is a total information function such that \( f(x, q) \in V_q \) for each \( q \in Q \), and \( x \in U \). In consequence, each object \( x \) of \( U \) is described by a vector which describes \( x \) in terms of the evaluations of the attributes from \( Q \) – that is, it represents the available information about \( x \). The explicit manifestation of this definition is to specify a decision table. As Greco, et. al. note, the decision table contains the information relative to a set of objects (school units), described by a certain number of attributes (LSI measures).

The decision problem here unfolds in two steps. The first step is to explain historic decisions in terms of the unique circumstances in existence at the time the decision was made. This is equivalent to the classification problem of MCDA. The second step is to provide a guide on how to make a current (or future) decision under specific circumstances. This latter step is equivalent to providing a sorted set of attributes that are known to best influence changes in school rank within dimension.

For the process of engineering an objective school classification system the rough sets approach to MCDA offers an alternative to classical MCDA. For example, the rough sets approach permits an extension of the decision-problem to include preferential information that is able to interrogate the relevance of attributes as well as their interaction effects. Additionally, and unlike the classical approach, rough sets can include fuzzy evaluations, missing values, and cardinal scales. The dual ANN based approach proposed in this research has many of these same features. As such, it represents a logical empirical approach to the MCDA problem of engineering a predictive school classification system.

2.1 Kohonen Self Organizing Map for Classification

A self-organizing map (SOM) is a neural network topology that implements a \( k \)-means cluster algorithm. The Kohonen SOM (K-SOM) has a supplementary property that permits its neuron layer to organize in a manner that results in a map of any input space to which it is exposed (see Kohonen [10] for theory and examples). As shown in Figure 1, the output of the K-SOM is a 2D map of ordered categories where each category (feature) occupies a space proportional to its included component’s frequency of occurrence. Stated differently, more frequent patterns occupy a greater area at the expense of the less frequent patterns. Research reports that SOMs produce computationally fast solutions that are extremely useful in classification when applied to
large high-dimensional data sets (Craven and Shavlik [11], Lu, et al. [12], Kaski & Kohonen [13], Kaski [14]).

The K-SOM implementation of this algorithm is particularly useful when the researcher does not know precisely what to look for as features within large unstructured databases. This fact is of particular importance to the classification of Rhode Island public high schools. The process is approached under the assumption that the precise identification of a defining feature set is unresolved.

2.2 The K4 Multiple Objective RBF ANN

In addition to classifying high schools, we would to understand how to objectively predict the assignment of individual schools to a pre-defined category. To this end we seek to predict (sort) the probability of category assignment for each individual school. To accomplish this task we employ an ANN to produce an analytic approximation of the input-output mappings generated by the noisy education data stream which may be described as,

\[
\{[x(k), y(k)]: ([\mathbb{R}^n, \mathbb{R}])\}_{k=1}^\infty
\]

where, \(x(k)\) is the input vector, or features, for entity \(k\), \(y(k)\) is the output for entity \(k\), and \(n\) is the dimension of the input space. The data is drawn from the noisy set described as

\[
\{[x(k), y(k) = f(x(k) + \varepsilon(k))]: k=1\}^\infty
\]

A shown in figure 2, the traditional RBF ANN topology is defined by three layers: input, hidden and output.
The input layer has no particular calculating power; its primary function is to distribute the information to the hidden layer. The hidden, or middle, layer embraces the computing units, or, hidden nodes. Each hidden node is defined by a center. The center is a parameter vector of the same dimension as the input data vector, \( x \), and calculates the Euclidean distance between the center and the network input vector defined by \( ||x(k) - c_j(k)|| \). The results are passed through a nonlinear activation function, \( \phi_j(k) \), to produce output from the hidden nodes. A popular choice of the activation function is the Gaussian basis function which is defined as

\[
\phi_j(k) = \exp \left( \frac{||x(k) - c_j(k)||^2}{\sigma_j^2} \right), \quad j = 1 \ldots n_h
\]  

(3)

where \( \sigma_j \) a positive scalar is called a width and \( n_h \) is the number of centers. The output layer is a linear combiner with the \( i \)th output of the network model being a weighted sum of the hidden node outputs,

\[
\hat{y}_i(k) = \sum_{j=1}^{n_h} \phi_j(k)w_{ji}, \quad i = 1 \ldots p
\]

(4)

where \( w \) are output layer weights, \( p \) is the number of outputs and \( \hat{y} \) is the network output to estimate the target \( y \) (for details of this description, see [15]). For additional discussion on the foundations of RBF ANNs, see D.S. Broomhead and D. Lowe [16], H. Lohinger [17], Parthasarathy and Narendra [18] as well as Sanner and Slotine [19].

5 The RBF ANN is a type of nonlinear regression model where it is possible to estimate the output layer weights by any of methods normally used in nonlinear least squares or maximum likelihood. However, we note that if every observation was used as an RBF center this would result in a vastly over-parameterized model.
The K4 extended RBF ANN was designed to consider multiple objectives within a Bayesian framework. To achieve the dual objectives of smoothness and accuracy, Kajiji augmented the Tikhonov regularization equation [20] to estimate the smoothing weight decay parameter, $v$, within a closed-form solution rather than the computationally burdensome iterative approach that had been in use (for a discussion on iterative solutions to this estimation see Hoerl and Kennard [21] and Hemmerle [22] and Orr [23-24]). By adding a weight penalty term to the SSE optimization objective, the modified SSE is restated as the following cost function:

$$C = \sum_{i=1}^{p} \left( \tilde{y}_i - f \left( x_i \right) \right)^2 + \sum_{j=1}^{m} v_j w_j^2,$$

(5)

where $v_j$ are the regularization parameters or weight decay parameters (see Crouse’s [25] Bayesian enhancement to optimal ridge regression). Under this specification the function to be minimized is stated as:

$$C = \frac{\text{argmin}}{v} \left( \sum_{i=1}^{p} \left( y_i - f \left( x_i \mid \bar{v} \right) \right)^2 + \sum_{j=1}^{m} v_j w_j^2 \right)$$

(6)

Taken together, the aforementioned extensions embraced by the K4 RBF ANN allow the dual-objective algorithm to directly attack the twin evils that deter efficient ANN modeling: the “curse” of dimensionality (multicollinearity or over-parameterization) and inflated residual sum of squares (inefficient weight decay). These enhancements have also proven beneficial when applying the K4 RBF ANN to both small and large data structures.

3. Model

Comprising only 1,545 square miles, the State of Rhode Island is the smallest U.S. state by geographical region. Despite its small geographic size, in AY2004-05 the state was divided into 39 cities and towns with 36 unique school districts plus 8 independent charter schools and 4 state-operated schools, for a total of 48 local education agencies. The research study examines the population of high schools within all school agencies. Although small for generalized statistical inference, the population of high schools within the state of Rhode Island totals 54.

3.1 Historical Context

Beginning with AY 1997-98 Rhode Island public schools administered the High Performance Learning Community Assessment: School Improvement Self-Study Survey (alternatively referred to as HiPlaces [26]) to all students, staff, parents, and administrators. Unfortunately, missing data is an issue as budget constraints negated the administration of the survey in AY2000-01 and AY2008-09. Despite budget woes, Rhode Island has been a forerunner in collecting this data and reducing it to meaningful information in the form of the 9x5 Model developed by Felner, et al. (http://www.ncpe.uri.edu/model/). The design assumption of the 9x5 Model is geared to provide school policymakers with information to implement and refine their efforts to ensure that all children achieve at high levels. In brief, the 9x5 model caches individual survey questions into forty-five subjective categories.
3.2 NCLB and the Development of the LSI-Styled Data Tables

The arrival of the NCLB with its heavy emphasis on scientific research and test-based proficiency rendered the 9x5 Model, and its focus to improve schools from within, burdensome. In an attempt to reconcile a continuation of the time series from AY 1997-1998 to current period, RIDE collaborated with NCPE to develop a smaller more manageable set of contextual school indicators. These indicators were labeled the LSI System. Three of the indicators are based on the 9x5 Model. Two additional indicators, Time Lost from School, and Graduation Rate were newly incorporated in the LSI System. (Note: Graduation Rate is not included in the LSI System for elementary and middle school levels). As stated earlier, the intent of the LSI System, see Figure 3, is to provide the information needed by local school decision-makers to develop and implement policies on teaching practice, school organizational structure and school culture in manner that produces annual improvement.

Alternatives to the LSI system have been proposed elsewhere in the literature. Daggett [27] collapsed his seven meta-analyses conducted over hundreds of subjects to ten central findings that school administrators could use to develop school reform. As we demonstrate in Table 1, the LSI’s index score approach also has a visible overlap with Daggett’s ten findings. Thus although the approaches and vocabulary may differ, both methods can be represented by a data table of information that is readily attacked by one or more MCDA methods.
Table 1: Equivalence of LSI Categories and Daggett’s 10 Focal Points

<table>
<thead>
<tr>
<th>LSI Categories</th>
<th>Daggett’s 10 Focal Points for Reform Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Climate (Q4)</td>
<td>Create a Culture</td>
</tr>
<tr>
<td>All collected data is analyzed at the school, district, and state level and distributed.</td>
<td>Use Data</td>
</tr>
<tr>
<td>Instruction (Q1, Q2, Q3, Q4)</td>
<td>Framework to Organize Curriculum</td>
</tr>
<tr>
<td>Parental Involvement (Q1, Q3); Attendance and Graduation</td>
<td>Accountability</td>
</tr>
<tr>
<td>School Climate (Q1), Parental Involvement (Q5)</td>
<td>High Expectation</td>
</tr>
<tr>
<td>Parental Involvement (Q3)</td>
<td>Continuous Improvement</td>
</tr>
<tr>
<td>Instruction (Q1, Q2, Q3)</td>
<td>Professional Development</td>
</tr>
<tr>
<td>Parental Involvement (Q1, Q2, Q3, Q4)</td>
<td>Parent &amp; Community Involvement</td>
</tr>
<tr>
<td>School Climate (Q3, Q4)</td>
<td>Safe and Orderly Schools</td>
</tr>
<tr>
<td>Instruction (Q3, Q4)</td>
<td>Effective Leadership Development</td>
</tr>
</tbody>
</table>

3.3 Data

Panel data representing the population of high-schools in Rhode Island for AY 2004-05 is obtained from two sources. The individual school classifications, attendance rates and graduation rates are obtained directly from reports published by RIDE. The HiPlaces based LSIs were obtained from the State Report Card [6]. The data table created has five predictor variables and one target variable – the prior year classification of each high school. Missing LSI values were replaced by the bootstrap method using data from previous years.

The predictor variables are described as follows. The proxy for the first LSI component, Time Lost from School is captured by including two observed measurements: Attendance rate and Graduation Rate. The attendance rate simply measures when a child is in school versus not in school. If the child has an out-of-school suspension the child is considered absent (or not attending school). The three survey-based LSIs are: Parental Involvement, Instruction, and School Climate. For discussion and calculation details on the LSI see: [http://www.infoworks.ride.uri.edu/2005/techbulletin](http://www.infoworks.ride.uri.edu/2005/techbulletin) . Variance stabilization for LSI predictor variables is achieved by natural logarithmic transformation [28].
4. Modeling Results

This section of the paper presents the results of the MCDA modeling. Section 4.1 focuses on the classification of schools from the LSI attributes. Section 4.2 presents the results of computing posterior classification probabilities via the Softmax transformation [29-31]. Section 4.3 discusses the results of applying the K4-RBF ANN to predict and sort R.I. schools into homogeneous performance groups.

4.1 Efficient Performance-Based Classification

In this section of the study, we apply the K-SOM method of classification to produce a three cluster mapping. Unlike the parametric-based methods (e.g., principal component analysis), the K-SOM method does not impose any distributional assumptions on the data components. Our objective is to employ this non-parametric visualization method to locate three latent clusters that correspond to the traditional performance classifications for the public high schools. Specifically, the model solved is:

\[ OC_i = f(P_1, P_2, ..., P_k), \]  

(7)

where \( OC_i \) is the school classification as defined by RIDE for \( i-th \) high school; \( P_1 \) is the \( \ln(\text{Attendance rate}) \), \( P_2 \) is \( \ln(\text{Graduation rate}) \), \( P_3 \) is \( \ln(\text{School Climate LSI}) \), \( P_4 \) is \( \ln(\text{Parental Involvement LSI}) \); and, \( P_5 \) is \( \ln(\text{Instruction LSI}) \). The result of applying the K-SOM algorithm to the classification model (equation 7) is a feature map with three latent but identifiable clusters as displayed in figure 4. Using the dependent axis scale, cluster formations at axis values 1, 2, and 3, are clearly evident.

Figure 4: K-SOM Clustering of School Achievement
Upon review of the detailed feature map as presented in Figure 4, the school classification scheme is restated to compare the actual associations against those generated by application of the K-SOM method. These comparative results are presented in Table 2 with the following interpretation. Following the presentation of the school name, the next column presents the actual classification provided by RIDE for AY2004-05. This is followed by the objectively determined classification gleaned from K-SOM feature map. Out of a total of 54 high schools, 12 were reclassified by the K-SOM procedure (approximately 22 percent). Of these 12 school reclassifications, 6 had an upward adjustment and, 6 had a downward adjustment as shown in the next column. It is noted that Rhode Island School for the Deaf, moved by more than one classification level. This is caused by a very low student to teacher ratio (33 teachers for 105 students) creating a bias towards the Instruction LSI. Further, unlike what one might expect there is a very low response rate from the parents of this school as compared with the state values (19% v/s 36%) further enhancing the bias towards the Instruction LSI. The last column shows the actual classification for AY2005-06 [32].

### Table 2: Re-Classification Table

<table>
<thead>
<tr>
<th>School Name</th>
<th>RIDE 2005 Performance Classification</th>
<th>K-SOM Performance Classification</th>
<th>K-SOM Predicted Direction</th>
<th>RIDE 2006 Performance Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burrillville High School</td>
<td>Moderately Performing</td>
<td>High Performing</td>
<td>+</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>NE Laborers’ Career Academy</td>
<td>Moderately Performing</td>
<td>In Need of Improvement</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>Rogers High School</td>
<td>Moderately Performing</td>
<td>In Need of Improvement</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>North Kingstown Senior High School</td>
<td>High Performing</td>
<td>Moderately Performing</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>North Providence High School</td>
<td>Moderately Performing</td>
<td>In Need of Improvement</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>William E. Tolman Senior High School</td>
<td>In Need of Improvement</td>
<td>Moderately Performing</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Shea Senior High School</td>
<td>In Need of Improvement</td>
<td>Moderately Performing</td>
<td>+</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>William B. Cooley/Health &amp; Science</td>
<td>Moderately Performing</td>
<td>In Need of Improvement</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>West Warwick Senior High School</td>
<td>In Need of Improvement</td>
<td>Moderately Performing</td>
<td>+</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>Blackstone Academy Charter School</td>
<td>Moderately Performing</td>
<td>In Need of Improvement</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>RI School for the Deaf *</td>
<td>In Need of Improvement</td>
<td>High Performing</td>
<td>+</td>
<td>Not classified</td>
</tr>
<tr>
<td>Metropolitan Regional Career &amp; Tech</td>
<td>Moderately Performing</td>
<td>High Performing</td>
<td>+</td>
<td>No change from 2005</td>
</tr>
</tbody>
</table>

* Needs further investigation. Their Survey based LSI information is very sparse.

### 4.2 The Prediction of Posterior Classification Probabilities

In this section we invoke an ANN model to predict classification probabilities for the purpose of sorting into pre-defined classification groups. To obtain the probability estimates, the K4 RBF
ANN is solved using the softmax activation function. This activation function forces the outputs to lie between zero and one and to sum to one. Specifically, the softmax output \( p_i \) is defined by equation 8, where the observed classification for the \( i \)th school, \( OC_i \), is transformed according to

\[
p_i = \frac{e^{OC_i}}{\sum_j e^{OC_j}}. \tag{8}
\]

The transformation produces \( p_i \) values of 0.090, 0.245, and 0.665, for categories 1, 2, and 3, respectively.

To examine the empirical relationship between the probability of group classification and the LSI system, we fit the cross section of data to a double-logarithmic functional transformation. Of course, the most well known production estimating equation of this type is the well-researched Cobb-Douglas (CD) production model [33]. The restrictive characteristics of the CD production function are known to be unrealistic for most production functions. Additionally, the CD form assumes that the dependent variable is a quantity of output that is formed from the use of factor inputs. Our objective here, however, is to obtain an explicit mapping of the target variable against the LSI predictor variables. The RBF ANN is known for its mapping efficiencies. Further, as developed in section 2 above, the K4 RBF ANN overcomes the curse of dimensionality that plagues many mapping algorithms. Hence, the final weights produced upon solving the K4 RBF ANN are treated as quasi-elasticity estimates. Specifically, we refer to the extracted weights from this ANN as K4-elasticities.

Let us denote \( S = \{s_1, ..., s_m\} \) as the set of schools, evaluated on the \( n \) criteria \( g_1, ..., g_n \), for \( n = 1...5 \), to be sorted into a predefined set of ordered categories \( \{C_1, ..., C_h\} \) where \( C_1 \) is the worst category and \( C_h \) is the highest (best) category. Subsequently, we execute the following production-theoretic model to test the predictive ability of the Kajiji-4 RBF ANN methods:

\[
p_i = f(P_1, P_2, ..., P_5), \tag{9}
\]

where \( p_i \) is the softmax classification probability for high school \( i \); \( P_1 \) is the ln(attendance rate), \( P_2 \) is ln(graduation rate), \( P_3 \) is ln(School Climate LSI), \( P_4 \) is ln(Parental Involvement LSI); and, \( P_5 \) is ln(Instruction LSI). That is, the economic functional form is:

\[
\ln p = \ln P_1 + \ln P_2 + \ln P_3 + \ln P_4 + \ln P_5 + e \tag{10}
\]

Table 3 presents the performance significance of alternative K4 RBF ANN algorithmic settings. That is, each column represents a unique data transformation prior to execution of the activation based ANN algorithm. For the purposes of this analysis we limit the discussion to the results produced by the data transformation STD:1 (standardized values). Compared with the alternative data transformations, this one yielded the smallest (most negative) AIC and Schwarz (BIC) coefficients. Additionally, both the Theil “U” (0.04) and MAPE measure (19.46) yield comparative low values. Not surprisingly, this solution also had the highest R-square at a reported value of 99.59 percent.
Table 3: Comparative Results of Alternative RBF Solutions

<table>
<thead>
<tr>
<th>Transformations</th>
<th>Norm:2</th>
<th>Norm:1</th>
<th>STD:2</th>
<th>STD:1</th>
<th>Uniform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computed Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambda</td>
<td>0.016</td>
<td>0.008</td>
<td>0.027</td>
<td>11.045</td>
<td>0.336</td>
</tr>
<tr>
<td>Actual Error</td>
<td>4.46E-02</td>
<td>2.26E-02</td>
<td>4.41E-02</td>
<td>7.76E-01</td>
<td>9.58E-02</td>
</tr>
<tr>
<td>Training Error</td>
<td>3.99E-02</td>
<td>4.66E-02</td>
<td>7.17E-02</td>
<td>5.66E-04</td>
<td>1.55E-02</td>
</tr>
<tr>
<td>Validation Error</td>
<td>5.59E-02</td>
<td>5.99E-02</td>
<td>1.09E-01</td>
<td>1.40E-03</td>
<td>2.01E-02</td>
</tr>
<tr>
<td>Fitness Error</td>
<td>4.79E-02</td>
<td>5.33E-02</td>
<td>9.07E-02</td>
<td>9.86E-04</td>
<td>1.78E-02</td>
</tr>
<tr>
<td>Performance Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direction</td>
<td>0.945</td>
<td>0.964</td>
<td>0.582</td>
<td>0.964</td>
<td>0.600</td>
</tr>
<tr>
<td>TDPM</td>
<td>0.013</td>
<td>0.014</td>
<td>0.023</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>R-Square</td>
<td>46.96%</td>
<td>36.61%</td>
<td>74.93%</td>
<td>99.59%</td>
<td>88.74%</td>
</tr>
<tr>
<td>Schwarz</td>
<td>-144.066</td>
<td>-138.357</td>
<td>-109.655</td>
<td>-353.804</td>
<td>-197.429</td>
</tr>
<tr>
<td>Theil</td>
<td>0.33</td>
<td>0.35</td>
<td>0.53</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>MAPE</td>
<td>127.68</td>
<td>126.90</td>
<td>135.96</td>
<td>19.46</td>
<td>49.70</td>
</tr>
<tr>
<td>Model Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training (N)</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Training (%)</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Min/Max/SD</td>
<td>0/1</td>
<td>0% / 1%</td>
<td>n/a</td>
<td>n/a</td>
<td>0/1/1.00</td>
</tr>
<tr>
<td>Radius</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Algorithmic Settings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>K4</td>
<td>K4</td>
<td>K4</td>
<td>K4</td>
<td>K4</td>
</tr>
<tr>
<td>Error Min. Rule</td>
<td>GCV</td>
<td>GCV</td>
<td>GCV</td>
<td>GCV</td>
<td>GCV</td>
</tr>
<tr>
<td>Transfer Function</td>
<td>Gaussian</td>
<td>Gaussian</td>
<td>Gaussian</td>
<td>Gaussian</td>
<td>Gaussian</td>
</tr>
</tbody>
</table>

Table 4 presents the solution weights obtained from solving equation 7 with STD:1 transformation. As noted, the RBF weights are like regression parameters obtained from a nonlinear regression. For this exercise, as K4-elasticity estimates, it is the sign and relative magnitude of the weights that is of importance to school policymakers. For example, the weight coefficient for the Attendance LSI is 0.385. From this “quasi” elasticity, policymakers could infer that a 10 percent increase in high school attendance explains a 3.8 percent contribution to the probability estimate. This finding is supported by the production estimates of Wilson [34] who she found extensive evidence that the utility-maximizing individual will choose schooling (attendance) in response to both the economic returns to schooling as well as the utility derived from schooling. Her study did caution that the amount of education in which an individual invests is conditional on the expected outcomes.

Table 4: Weights from Alternative RBF Solutions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Norm:2</th>
<th>Norm:1</th>
<th>STD:2</th>
<th>STD:1</th>
<th>Uniform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>1.592</td>
<td>1.866</td>
<td>-0.448</td>
<td>0.385</td>
<td>0.037</td>
</tr>
<tr>
<td>Graduation (%)</td>
<td>-1.835</td>
<td>-2.115</td>
<td>0.015</td>
<td>0.095</td>
<td>0.289</td>
</tr>
<tr>
<td>LSI School Climate</td>
<td>0.621</td>
<td>0.631</td>
<td>1.338</td>
<td>0.324</td>
<td>0.116</td>
</tr>
<tr>
<td>LSI Instruction</td>
<td>1.583</td>
<td>1.990</td>
<td>0.024</td>
<td>0.286</td>
<td>0.045</td>
</tr>
<tr>
<td>LSI Parental Involvement</td>
<td>-1.976</td>
<td>-0.973</td>
<td>-1.230</td>
<td>0.233</td>
<td>0.341</td>
</tr>
</tbody>
</table>

For the other three LSI’s (School Climate, Instruction, and Parent Involvement) the comparative contributions are also positive albeit somewhat smaller in comparative magnitude. Of these three, the School Climate LSI produced a K4-elasticity of 0.324. With a 10-percent change in high school climate, there is an expectation of changing the associated classification probability...
by 3.24 percent. These results also buttress the early findings of Wilson. School climate and the expected utility of schooling are closely related. Lazear [35] provides convincing arguments for a focus on class size, noting that discipline (a climate factor) is a substitute for class size. Dee [36] provides new detailed findings that when school district administrators use available resource effectively then, *ceteris paribus*, the money that is spent on instruction does matter. Recently European-based findings provided by Dreher, et. al. [37] argue for targeted aid to education to improve national educational outcomes. We note that the K4-elasticity on the *Instruction* dimension is within specification with an estimate of 0.286. Taken together these findings point to the importance of how school climate and instruction initiatives come together to affect the potential for changing classification probabilities. The smallest weight for probabilistic prediction is attached to *Graduation Rate*. However, clearly graduation rate is a contributing factor to overall school climate, instruction and attendance.

The *Parent Involvement* LSI has an interesting K4-elasticity. Achieving a measure of 0.233 in the targeted analysis, we note the negative value produced by three alternative data transformation methods. The literature and history of the *Parent Involvement* dimension has long suggested a positive, but small, correlation with school performance (see, Izzo, et. al., [38]). The findings here corroborate this history; and, the conflicting signs add to questions about the dimensions overall contribution under different modeling assumptions. The findings presented here corroborate the extant literature; and, the observed conflicting parameter signs are consistent with deterioration of parental support hypothesis of Houtenville, et. al. [39] when school resources increase over time.

Finally, in Table 5, we present the classification derived from the predicted probability of the RBF solution. These values were compared with the original classification presented in Table 2. It is important to note that nearly all schools maintained their *K-SOM* classification. The schools that did have a different classification are shown Table 6. Although the list contains only five schools, the post-probabilistic classification returned two schools to their pre-analytic classification. Of the three new schools, all experienced a performance over-rating. That is, probabilistic classification reduced what appear to be excessively high ratings to a more moderate level.

<table>
<thead>
<tr>
<th>Min Value</th>
<th>Max Value</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.148</td>
<td>1</td>
</tr>
<tr>
<td>0.149</td>
<td>0.403</td>
<td>2</td>
</tr>
<tr>
<td>0.403</td>
<td>Max</td>
<td>3</td>
</tr>
</tbody>
</table>
### Table 6: Schools with Different K-SOM and RBF Performance Classification

<table>
<thead>
<tr>
<th>School Name</th>
<th>RIDE 2005 Performance Classification</th>
<th>K-SOM Performance Classification</th>
<th>RBF Performance Classification</th>
<th>RBF Direction</th>
<th>RIDE 2006 Performance Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middletown High School</td>
<td>High Performing</td>
<td>High Performing</td>
<td>Moderately Performing</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>North Providence High School</td>
<td>Moderately Performing</td>
<td>In Need of Improvement</td>
<td>Moderately Performing</td>
<td>Same as RIDE 2005</td>
<td></td>
</tr>
<tr>
<td>Westerly High School</td>
<td>High Performing</td>
<td>High Performing</td>
<td>Moderately Performing</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>Wm M. Davies Jr. Career &amp; Tech</td>
<td>High Performing</td>
<td>High Performing</td>
<td>Moderately Performing</td>
<td>-</td>
<td>No change from 2005</td>
</tr>
<tr>
<td>Metropolitan Regional Career &amp; Tech</td>
<td>Moderately Performing</td>
<td>High Performing</td>
<td>Moderately Performing</td>
<td>Same as RIDE 2005</td>
<td></td>
</tr>
</tbody>
</table>

### 5. Summary and Conclusions

The purpose of this paper was two-fold. First, using $K$-SOM methodology analytic results produced objective and efficient classification based on the Rhode Island target LSI measures. Second, using a combination of the softmax activation function and the K4 RBF ANN with a standardized value transformation of all inputs, an efficient mapping of classifications probabilities was estimated. These predictive probabilities proved to be realistic in ex-post classification and re-classification. This finding adds to the support for the use of LSI indices, or by inference close counterparts, to serve as accurate predictors of school performance. This finding clearly suggests that school administrators should focus on implementing policies and programs that would help increase their school’s LSI scores focusing first on school attendance and the internal climate of the school.
6. References


Founded in 1892, the University of Rhode Island is one of eight land, urban, and sea grant universities in the United States. The 1,200-acre rural campus is less than ten miles from Narragansett Bay and highlights its traditions of natural resource, marine and urban related research. There are over 14,000 undergraduate and graduate students enrolled in seven degree-granting colleges representing 48 states and the District of Columbia. More than 500 international students represent 59 different countries. Eighteen percent of the freshman class graduated in the top ten percent of their high school classes. The teaching and research faculty numbers over 600 and the University offers 101 undergraduate programs and 86 advanced degree programs. URI students have received Rhodes, Fulbright, Truman, Goldwater, and Udall scholarships. There are over 80,000 active alumnae.

The University of Rhode Island started to offer undergraduate business administration courses in 1923. In 1962, the MBA program was introduced and the PhD program began in the mid 1980s. The College of Business Administration is accredited by The AACSB International - The Association to Advance Collegiate Schools of Business in 1969. The College of Business enrolls over 1400 undergraduate students and more than 300 graduate students.

**Mission**

Our responsibility is to provide strong academic programs that instill excellence, confidence and strong leadership skills in our graduates. Our aim is to (1) promote critical and independent thinking, (2) foster personal responsibility and (3) develop students whose performance and commitment mark them as leaders contributing to the business community and society. The College will serve as a center for business scholarship, creative research and outreach activities to the citizens and institutions of the State of Rhode Island as well as the regional, national and international communities.

The creation of this working paper series has been funded by an endowment established by William A. Orme, URI College of Business Administration, Class of 1949 and former head of the General Electric Foundation. This working paper series is intended to permit faculty members to obtain feedback on research activities before the research is submitted to academic and professional journals and professional associations for presentations.

An award is presented annually for the most outstanding paper submitted.