An intelligent system to predict academic performance based on different factors during adolescence

A. T. M. Shakil Ahamed, Navid Tanzeem Mahmood and Rashedur M Rahman

Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

ABSTRACT
Students need to have an effective education to take advantage of all the latest tools available. Even with a proper education, they are failing to reap its benefits; reasons involve social, economic and psychological factors a student faces during their adolescence. Our research is directed towards this particular problem of educational effectiveness. We have surveyed a large number of students across different districts in Bangladesh. Pre-processing was done thoroughly; the use of data balancing, dimensionality reduction, discretization and normalization in combinations has allowed us to derive the best model that could predict the academic performance based on different factors during the adolescence.

ARTICLE HISTORY
Received 1 April 2016
Accepted 15 March 2017

KEYWORDS
Educational data mining; predictive model; machine learning

Introduction
Analysing failure rate of college students is a new area for research in Bangladesh. From 1991 to 2001, dropout rates have been reduced from 59.3% to 35% in primary and secondary level schools by setting policies; however, problems still persist at undergraduate level education (Islam & Pavel, 2001). We have aimed to investigate socioeconomic and psychological factors that have influential impact on academic attainment of students. Ineffective education can be fixed by adding critical policies; finding these policies is what our study aims to do.

Predicting the Higher Secondary grade can provide a proper idea of what a student will be able to achieve in life. Students undergo adolescence before their HSC exam and during this phase they are more susceptible to their surroundings. If we could ensure that they receive the proper care and guidance during this time, an effective education will help them pursue a successful career later.

Educational institutions are the main source of data along with some coaching centres. Our target students are ones who have just passed 12th grade. We have pre-processed the raw data by applying data balancing, normalization and optimal equal width binning (OWB) for discretization in order to design the predictive model. We have also applied three different dimensionality reduction algorithms that reduce the number of attributes that are used during training. Since we have used several pre-processing techniques, we
have comprised our comparative study with the combination of different pre-processing models to derive the best data model for our system.

**Related work**

Neural network has been used by Gedeon and Turner (1993) to perform a comparative study across several types of neural networks, including feed forward and back propagation that had been set up to forecast students’ final grades. Another system has been developed by Wang and Mitrovic (2002) wherein they used neural networks’ back propagation and feedforward techniques to predict the number of mistakes a student would make. The likeliness of being accepted by a university has also been predicted by Oladokun, Adebjano, and Charles-owaba (2008) using multilayer perceptron topology. Jishan, Rashu, Mahmood, Billah, and Rahman (2015) worked on prediction of educational performance using the optimal width binning technique for discretization of continuous valued attribute. Yadav, Bharadwaj, and Pal (2012) have used the attendance of the students, seminar, class test and assignment marks as attributes for predicting students’ final grade of a semester. Ayers, Nugent, and Dean (2009) used model-based clustering, $K$-means and other clustering algorithms to understand and visualize the skill level of an individual student and tried to group the students according to their skill sets. Pal and Pal (2013) used several classification techniques to cluster such students who need special care such as counselling from the teacher.

In our earlier study (Ahamed, Mahmood, and Rahman, 2016), we designed a predictive model where we have applied several techniques, e.g. one-r, gain ratio, to identify the relevant attributes which are highly correlated with the target class in order to reduce the classification error and to address the overfitting problem in prediction. We have used multi-interval discretization as the pre-processing technique to discretize our data set and then we have applied Random forest and Naïve Bays to predict the academic attainment of students. As data were unbalanced, in a later study (Ahamed, Mahmood, and Rahman, 2017) we applied an advanced balancing technique and used principal component analysis (PCA) on balanced data and report performance of classifiers. Due to the continuous nature of our target class, we have had to discretize it before using machine learning techniques. We have extended our work in this study by scrutinizing a comparative study of different dimensionality techniques such as PCA, self-organizing map (SOM) and generalized Hebbian algorithm (GHA). We inspect our key finding thoroughly by appending and discovering key rules for three categories, which are based on students’ academic performance, (i) Good [HSC 4–5], (ii) Average [HSC 3–3.75] and (iii) Bad [HSC 1–2.75].

**Designing the survey questionnaire and the collection of data**

Designing a predictive model requires a data set that has the essential attributes to predict future academic performance. After careful consultation with relevant experts, we have managed to design a questionnaire that contains most relevant attributes regarding academic attainments of the students. Our survey consists of 38 questions amongst which 5 are psychological questions that increase the likeliness of students answering truthfully. We have interviewed students from different economic statuses, cultures and locations to maintain the proper sampling and versatility in the data set. We have surveyed 423 students, amongst whom 8 were excluded from the data set due to missing information.
**Socioeconomic/demographic factors**

*Sex and age* have a significant impact on their academic achievement (Aripin, Mahmood, Rohaizad, Yeop, & Anuar, 2008).

*Father’s and mother’s education* denotes the social status of a family and is a strong predictor of academic performance (Coleman et al., 1966).

*Father’s and mother’s jobs* impact academics of the students (Usaini & Abubakar, 2015).

*Average monthly family income* indicates the economic status of that family, which has strong correlation with academic performance (Adler et al., 1994).

*Place of education* has a notable effect on academic outcomes (Cheers, 1990).

*Private tuition* significantly augmented the academic performance (Atta et al., 2011).

**Psychological factors**

*Family size* denotes the level of involvement of parents in education (Odok, 2013).

*Age difference with immediate sibling* birth spacing between children might have a long-term effect on academic performance (Buckles & Munnich, 2012).

*Parents’ involvement in education*: Christian, Morrison, and Bryant (1998) have observed that responsive parenting in education has a notable impact on academic achievements.

*Parents’ marital status*: Hatzichristou (1993) has notified that marital status of parents has a psychological effect on students, leading to an impact on academic outcomes amongst adolescents.

*Absence/death of parent(s)* has a sobering effect on mental health and causes impacts on academic outcomes (Jeynes, 2002).

*Involvement in physical activity*: Keeley and Fox (2009) stated that physical activities help to augment cognitive performance, leading to better academic achievement amongst students.

*Motivating teacher* helps students to be confident in learning, which plays a significant role in academic achievements (Adeyemo, Oladipipo, & Omisore, 2013).

*Motivating parents* significantly pass their motivation on to the students, making them confident and helping them become a better academic performer (Aatta & Jamil, 2012).

*Academic ambition*: Schoon and Parsons (2002) have stated that ambition in education can help students to achieve academic attainments.

*Time spent on social media*: Flad (2010) has observed through a test amongst students that using social media network has an impact on academic performance.

*Romantic relation*: Halpern, Joyner, Udry, and Suchindran (2000) have indicated that romantic involvement is correlated to academic attainments.

*Time spent with friends*: a correlation between a student’s performance and peers’ influence has been proven by Kirk (2000) through specific studies.

*Alcohol/drugs consumption*: Aertgeerts and Buntinx (2002) have notified that alcohol consumption has an impact on academic outcomes.

*Smoking* hampers memory function and cognitive activeness, which significantly influences academic performance (Garrett, Dube, Trosclair, Caraballo, & Pechacek, 2011).

*Health status of participant*: Behrman and Lavy (1998) have noted that the statues of health have an impact on the cognitive performance of students.

*Health status of family members*: Parental involvement (Christian et al., 1998) in the education of a student is notably correlated with their health status.
Academic factors

Early childhood education has a long-term effect on further educational performance throughout a student’s life (Johnson, 1996).

Weekly study time: McFadden and Dart (1992) have observed through some studies that the amount of study time has an appreciable impact on academic achievement.

Previous failures: academic performance notably depends on previous failure(s) (Brunstein & Gollwitzer, 1996).

Number of absence(s): Kirby and McElroy (2003) state that a student’s attendance is a significant predictor of academic performance.

Extracurricular activities help students gather life skills, which are directly or indirectly correlated to academic attainments (Massoni, 2011).

Accessibility of the Internet helps students to connect with the latest information and can significantly broaden their knowledge, influencing academic performance (Ellore, Niranjan, & Brown, 2014).

SSC (Secondary School Certificate) 10th standard result: Previous academic record has an influence on future academic performance (Anderson, Benjamin, & Fuss, 1994).

HSC result: Result of HSC.

Designing the predictive model

Figure 1 shows the overview of the predictive model. The first few steps of pre-processing involve data balancing, followed by using three different dimensionality reduction techniques in a comparative study to derive the best performer(s) amongst them. Due to the continuous nature of our target class, we have had to discretize it. Lastly, normalization was used to ensure equal contribution of all features in the predictive model. After using all the pre-processing techniques, we have applied three different classification algorithms in a comparative study to predict the grade of higher secondary result.

Data balancing

SMOTEBost (Chawla, Lazarevic, Hall, & Bowyer, 2003) is an effective technique which includes the combination of SMOTE (Synthetic Minority Oversampling Technique) which fixes skewed data sets by balancing the minority and majority class and boosting the procedure to balance the data set. Figures 2 and 3 describe the data set before and after using SMOTEBost.

SMOTEBost requires a series of operations, $T$, to perform the task of balancing the data set. For each iteration, it calls a weak learner which is represented by a distribution $D_t$ and is altered by the particular training instances.

The distribution gets updated by assigning a higher weight to a wrong classification than a correct one. SMOTE is used in each round of boosting to oversample the minority class for every learner. In order to measure the weak hypothesis, $h_t$, the whole weighted training set is provided to the weak learner. At the end, a final hypothesis is computed from a combination of the different hypotheses.
Dimensionality reduction

In a data set, two or more different attributes could relate to the target attribute very similarly; so similarly in fact that one increases accuracy and the other creates noise. Sometimes it is observed that some attribute(s) have a very weak or even non-existent relationship with the target attribute. Classifier performance improves by removing unwarranted properties from the high-dimensional data set (Jimenez & Landgrebe, 1998). We have used three algorithms, namely, PCA, GHA and SOM.

Figure 1. An overview of our system.
Principal component analysis

PCA converts redundant dimensions into a set of attributes, which are correlated only to the target attribute using an orthogonal transformation (Jolliffe, 2002). The first attribute generated by PCA is meant to have the highest variance and each successive attribute is made to have the next highest variance. Given a data set $x$, the $k$th component can
be found by using the following formula:

\[ \hat{x}_k = x - \sum_{s=1}^{k-1} xW(s)W^T(s). \] (1)

PCA also needs to calculate the loading vector, \( W(k) \): the highest achievable variance using the following formula:

\[ W(k) = \arg \max \{ ||\hat{x}_kw||^2 \} = \arg \max \left\{ \frac{w^T\hat{x}_k\hat{x}_kw}{w^Tw} \right\}. \] (2)

For our PCA setup, we used a variance threshold; the accumulated variance of any attribute greater than a threshold of 0.9 gets removed.

**Self-organizing map**

This algorithm implements learning by competition; the use of a neighbourhood approach retains the topological characteristics of the data set. Euclidean distance up to all weight vectors is calculated when the data are passed through the network. The weight vector that matches the input neuron the best is known as best matching unit (BMU) (Kohonen & Honkela, 2007). The BMU’s weights and surrounding neurons in the SOM structure are moved to the input vector. The amount of change reduces with time and distance from the BMU inside the lattice. Equation (3) describes the relation between neuron \( v \) and weight vector \( W_v(s) \)

\[ W_v(S + 1) = W_v(S) + \Theta(u, v, s)a(S)(D(T) - W_v(S)), \] (3)

where \( S \) is the step index, \( t \) is the index into training sample, \( u \) is the index for BMU for \( D(t) \), \( a(S) \) is the learning coefficient, \( D(t) \) is the input vector, \( \Theta(u, v, s) \) is the neighbourhood function; it holds the distance between neurons \( u \) and \( v \) for step \( s \).

With a net size of 25, starting learning rate of 0.5 and adaptation radius of 20, SOM trains over 100 cycles from the data set. The learning rate and adaptation radius are reduced in each cycle until it reaches the finishing values of 0.01 and 1, respectively.

**Generalized Hebbian algorithm**

This particular algorithm is based on an acyclic and linear neural network which is used for unsupervised learning. The approach it uses involves just one layer of learning based on the weight of the neural node using the formula shown below. This weight depends on the correlation between the input attributes and output attribute (Gorrel, 2006)

\[ \Delta W_{ij} = \eta \left( Y_iX_j - Y_i \sum_{k=1}^{i} W_{kj}Y_k \right), \] (4)

where \( W_{ij} \) is the synaptic weight between the \( j \)th input neuron and the \( i \)th output neuron, \( X \) is the input vector, \( Y \) is the output vector and \( \eta \) is the learning rate.

A simple and forecastable balance between the effectiveness of learning and precision of convergence can be set using just one parameter. We have configured GHA to cap the maximum number of attributes to 15 over 30 iterations with a learning rate of 0.07.
Optimal equal width binning

OWB has been used to discretize the continuous data into specific intervals; having the same interval length, along with error minimization as stated in Kayah (2008) for our system, an optimal bin width can be obtained by dynamically iterating the searching process until we get the optimal value for the bin width. Using dynamic width values over several iterations and using the Naïve Bayes classifier for a comparison of accuracies, the best number of bins was 8 and the best bin width was 0.5 for our data set: $[-\infty,-1.5)$, $[-1.5,-2)$, $[-2,-2.5)$, $[-2.5,-3)$, $[-3,-3.5)$, $[-3.5,-4)$, $[-4,-4.5)$ and $[-4.5,\infty]$.

Normalization

Our data set contains attributes with different units and scale. Therefore, it requires normalization and in our study we have applied min–max normalization to bring attributes into a comparable range. According to Jayalakshmi and Santhakumaran (2011), the min-max method rescales attributes by changing the range of values to a new range which is rescaled from 0 to 1 or $-1$ to 1. The process in which this technique rescales often is implemented by the following formula:

$$x' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (5)$$

when $x_{\max} - x_{\min} = 0$ for a feature, it indicates a constant value in the data set for that particular feature. When a feature with constant value is found in the data set, it gets removed immediately as it does not carry any necessary information for the learner.

Learning algorithms

We have used stratified sampling and made a 70:30 split for training and testing. Stratified sampling retains class distribution properly for each split. This ensures proper validation after training.

Artificial neural network

Complex chemical substances are discharged from neurons to send information to other neurons using synapses to increase or decrease the electrical potential within the receiving neuron; when a certain electrical potential is reached, the neuron fires. This is the feature that the artificial neurons try to replicate. After setting up the network, the neurons are assigned numbers so that they are distinguishable. The artificial neuron has the following activation function:

$$a_i = \sum_{j=1}^{N} W_{ji} x_j + \theta_i, \quad (6)$$

where $x_j$ is the user input or output of a predecessor neuron, $\theta_i$ is the threshold and $W_{ji}$ is the weight. We have used the ‘feed forward’ neural network; the input of each layer is fed from the output generated by the previous layer. Similarly, the output of each layer is
considered to be used as the input of the subsequent layer. One of the constraints of the ‘feed forward’ architecture is that neurons are not enabled to connect to any neuron on the same or prior layer. The layers situated between input layer, the first layer and the output layer, the last layer are called hidden layers (Kar, n.d.).

**K-nearest neighbour**

In the K-nearest neighbour (K-NN), all the attributes and the target attribute are mapped into a vector space where every vector represents the N dimensions for each attribute. Finding the number of neighbours (K) is crucial because a very low number of K might lead to a noisy classification whereas a high number of K might lead to overfitting. We have used cross-validation to get an optimized value of K; the data set is divided into V folds of samples. For a fixed number K, we use K-NN for the classification and get the sum of the squared error for the vth fold. We have repeated the process for different numbers of K for the V segments. Then we got the optimized K value from the cross-validation having the lowest error (Mullin & Sukthankar, 2002). After getting the optimized K value, the Euclidean distance is used to measure the similarity metric using

\[ d_E(x, y) = \sum_{i=0}^{N} \sqrt{(x_i^2 - y_i^2)}, \]

where \( d_E \) is the Euclidean distance between the instances \( x_i \) and \( y_i \). After that, hamming distance is used to get the similarity metric which is based on the following equation:

\[ D_H = \sum_{i=1}^{k} |x_i - y_i|, \]

where \( D_H \) is the hamming distance between categorical instance \( x_i \) and \( y_i \). Then a majority vote approach is used amongst the k neighbours and assigns a classification to the target outcome (Alkhatib, Najadat, Hemeidi, & Ali Shatnawal, 2013).

The K-NN is used in our data set in the following ways:

1. Determining appropriate value of k to get the similar neighbours used to classify.
2. Computing the distance between the instances of training and testing samples.
3. The samples are then sorted relative to their distance.
4. Using majority vote of K-NNs, the result for HSC is predicted.

**Support vector machine**

Support vector machine (SVM) constructs hyper planes in a multidimensional space to perform its classification, which separates class levels into different cases. An iterative training algorithm is used in SVM in order to construct the optimal hyper plane, which leads to optimizing an error function by minimizing it iteratively. We have used c-SVM
to train our model in which the error function is characterized by (Min & Lee, 2005)

$$\frac{1}{2} w^T w + c \sum_{i=1}^{N} \xi_i$$

(9)

Subject to: $$y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \ldots, N,$$

where $$C$$ is the capacity function, $$w$$ are the vectors of coefficients, $$b$$ is the constant, $$\xi_i$$ is the parameter that handles non-separable data, $$i$$ is the number of training cases ($$N$$), $$y \in \pm 1$$ are the class labels, $$x_i$$ is the independent variable; kernel $$\phi$$ transforms data from the input to the feature space. As $$C$$ gets larger the error gets more penalized; it is crucial to choose $$C$$ very carefully to avoid getting the over-fitting problem. We have used dot kernel for designing our system, which can be characterized as follows:

$$K(X_i, X_j) = \begin{cases} X_i \cdot X_j & \text{linear}, \\ (\gamma X_i \cdot X_j + C)^4 & \text{polynomial}, \\ \exp(-\gamma |X_i - X_j|^2) & \text{RBF}, \\ \tanh(\gamma |X_i \cdot X_j| + C) & \text{sigmoid}. \end{cases}$$

(10)

In Equation (10), the kernel function used the dot product method to map input data point transforming to a feature space of high dimension.

### Implementation of the system

All the training and testing models we have developed have been built on RapidMiner; a software that allows proper integration of all the important tools we need for training and testing different predictive models. SMOTEBoost was implemented in WEKA; dimensionality reduction algorithms, namely, PCA, SOM and GHA, were all implemented in RapidMiner. We have used OWB to discretize the target class followed by normalization; both implemented in RapidMiner. After all the pre-processing is done, we have used three learning algorithms, namely, Artificial neural network (ANN), K-NN and SVM in RapidMiner.

### Result analysis

#### Key findings

In our earlier studies (Ahamed et al., 2016, 2017) we have reported some of our initial and general findings. In this research we have made a depth analysis by discovering rules for a specific group which is formed by different intervals of CGPA. In this way, the reasons for academic performance of that group of CGPA could be discovered. In most cases, students perform similarly in both HSC and SSC exams as shown:

- IF SSC = 1 THEN HSC = 1
- IF SSC = 2.250 THEN HSC = 2.500
- IF SSC = 3.750 THEN HSC = 4.000
- IF SSC = 5 THEN HSC = 5
When teachers or parents motivate students, they tend to perform better, especially when they have previous failures. Private tuition helps academically when a student has suffered from previous failures as shown:

\[
\text{IF PREVIOUS FAILURES} = 2 \\
| \text{IF MOTIVATING TEACHER} = \text{NO} \\
| | \text{IF PRIVATE TUITION} = \text{NO} \\
| | | \text{IF FAMILY SIZE} = \text{MORE THAN THREE} \quad \text{HSC} = 1.000 \\
| | | \text{IF FAMILY SIZE} = \text{LESS THAN THREE} \quad \text{HSC} = 3.250 \\
| | | \text{IF PRIVATE TUITION} = \text{YES} \quad \text{HSC} = 3.250 \\
| | \text{IF MOTIVATING TEACHER} = \text{YES} \quad \text{HSC} = 2.250 \\
\]

\[
\text{IF PREVIOUS FAILURES} = 3 \\
| \text{IF MOTIVATING TEACHER} = \text{NO} \\
| | \text{IF HANGING OUT WITH FRIENDS} = 2 \quad \text{HOURS} \quad \text{HSC} = 3.250 \\
| | \text{IF HANGING OUT WITH FRIENDS} = 3 \quad \text{HOURS} \quad \text{HSC} = 2.750 \\
| | \text{IF HANGING OUT WITH FRIENDS} = 4 \quad \text{HOURS} \quad \text{HSC} = 3 \\
| | \text{IF HANGING OUT WITH FRIENDS} = 5 \quad \text{HOURS} \\
| | | \text{IF PLACE OF EDUCATION} = \text{RURAL} \quad \text{HSC} = 3.500 \\
| | | \text{IF PLACE OF EDUCATION} = \text{URBAN} \quad \text{HSC} = 1 \\
| | \text{IF MOTIVATING TEACHER} = \text{YES} \quad \text{HSC} = 3.500 \\
\]

Parents’ marital status does seem to have an effect on the academic performance of students as shown below:

\[
\text{IF PARENT STATUS} = \text{LIVING APART} \quad \text{HSC} = 4 \\
\text{IF PARENT STATUS} = \text{LIVING TOGETHER} \quad \text{HSC} = 4.500 \\
\]

We observe once again that a family with less than or equal to three members offer students more attention and results in a better HSC result as shown:

\[
\text{IF PARENT STATUS} = \text{LIVING APART} \quad \text{HSC} = 3.750 \\
\text{IF PARENT STATUS} = \text{LIVING TOGETHER} \\
| \text{IF FAMILY SIZE} = \text{MORE THAN THREE} \quad \text{HSC} = 3.750 \\
| \text{IF FAMILY SIZE} = \text{LESS THAN THREE} \quad \text{HSC} = 4 \\
\]

Romantic relationship sometimes gives a significant boost in academic performance over SSC result because students aged 17–18 are more inclined to come to class for their romantic partner shown below:

\[
\text{IF SSC} = 2.500 \\
\]
IF ROMANTIC RELATION = NO THEN HSC = 2.500  
IF ROMANTIC RELATION = YES THEN HSC = 2.750

IF SSC = 2.750
| IF ROMANTIC RELATION = NO THEN HSC = 2.500
| IF ROMANTIC RELATION = YES
| IF AGE = 17 THEN HSC = 3
| IF AGE = 18 THEN HSC = 1.750

To further study the patterns generated by our decision tree we have split the tree based on three categories of students: (i) Good [HSC 4–5], (ii) Average [HSC 3–3.75] and (iii) Bad [HSC 1–2.75].

**Range: [1–2.75]**

Romantic relationship proved to be a complicated pattern; sometimes it was producing contradictory results. However, once we investigated further, we saw that it is highly based on parents’ marital status as shown:

IF PARENT STATUS = LIVING APART
| IF ROMANTIC RELATION = NO THEN HSC = 2.750
| IF ROMANTIC RELATION = YES THEN HSC = 1

IF PARENT STATUS = LIVING TOGETHER
| IF ROMANTIC RELATION = NO THEN HSC = 2.250
| IF ROMANTIC RELATION = YES THEN HSC = 2.750

The following subtree shows that a romantic relationship reduces academic performance for students who go to urban schools but academic performance improves if the student goes to a rural school as shown:

IF PLACE OF EDUCATION = RURAL
| IF ROMANTIC RELATION = NO THEN HSC = 2.250
| IF ROMANTIC RELATION = YES THEN HSC = 2.750

IF PLACE OF EDUCATION = URBAN
| IF ROMANTIC RELATION = NO THEN HSC = 2.500
| IF ROMANTIC RELATION = YES THEN HSC = 1

**Range: [3–3.75]**

Students who do not want higher studies are correlated with their weekly study time to predict their academic performance and we can see that the higher the weekly study time, the better the performance as shown:
IF HIGHER STUDY AMBITION = NO
| IF WEEKLY STUDY TIME = 1 HOUR THEN HSC = 3
| IF WEEKLY STUDY TIME = 2 HOURS THEN HSC = 3.250
IF HIGHER STUDY AMBITION = YES THEN HSC = 3.500

Range: [4–5]

In our study we have found that SSC result best reflects what can be expected of a student from his or her HSC exam as shown:

IF SSC = 3.500 THEN HSC = 4.000
IF SSC = 3.750 THEN HSC = 4.000
IF SSC = 5.000 THEN HSC = 5.000

Other patterns found through our study

(1) When a student spends time with friends a lot and he or she goes to an urban school, they perform worse compared to students who go to rural schools.
(2) When a student without private tuition is part of a family with more than three members, they fail to perform well academically.
(3) Romantic relationship’s effect on academics is dependent on unique factors:
   (a) If parents do not live together, students perform worse.
   (b) If parents live together, students perform better.
(4) There is a reduction in academic performance for students who go to urban schools.
(5) Academic performance improves if the student goes to a rural school.
(6) Students who want to pursue higher studies generally perform better than those who do not.
(7) Those who do not want to pursue higher studies require higher number of study hours per week to perform better.
(8) SSC results provide a good estimate of what is to be expected from the HSC exam.

Comparison of dimensionality reduction methods

We have compared the dimensionally reductions techniques PCA, SOM and GHA with each data model. Tables 1–4 contain the performances of these techniques.

Without any pre-processing

SVM with PCA and no dimensionality reduction scores accuracies of 77.88% and 76.50%, respectively shown in Table 1. K-NN performs the worst for this data model; the problem of overfitting plagues K-NN and causes 50.23% to be the highest achievable accuracy. ANN performs equally well for PCA and no dimensionality reduction with an accuracy of 70.05% for each.
As soon as SMOTEBoost is applied, accuracy for ANN increases by 1% for PCA and for GHA it boosts up by almost 8% as shown in Table 2. K-NN is again the worst performer here; the only significant accuracy boost comes from using SOM which is about 3%. For every other model K-NN has lower accuracy. SVM also has a reduced accuracy for PCA and no dimensionality reduction; reduction of about 1% and 7%, respectively. However, SVM has gained about a 10% higher accuracy for the data model pre-processed using GHA.

Table 1. Performance of the models without any pre-processing.

| Models      | Accuracy (%) | Precision (%) | Recall (%) | F measure (%) |
|-------------|--------------|---------------|------------|---------------|
| ANN + PCA   | 70.05        | 55.96         | 59.21      | 57.54         |
| ANN + SOM   | 32.72        | 11.57         | 20.43      | 14.77         |
| ANN + GHA   | 62.67        | 58.22         | 53.88      | 55.97         |
| ANN         | 70.05        | 61.64         | 50.30      | 55.40         |
| K-NN + PCA  | 30.35        | 16.32         | 19.25      | 17.66         |
| K-NN + SOM  | 27.19        | 25.04         | 27.01      | 25.99         |
| K-NN + GHA  | 19.35        | 18.01         | 18.32      | 18.16         |
| K-NN        | 50.23        | 54.60         | 50.36      | 52.39         |
| SVM + PCA   | 77.88        | 71.99         | 55.52      | 62.69         |
| SVM + SOM   | 36.87        | 20.62         | 24.39      | 22.35         |
| SVM + GHA   | 64.52        | 60.31         | 42.33      | 49.75         |
| SVM         | 76.50        | 71.00         | 58.46      | 64.12         |

Note: It can be observed that SOM offers the worst pre-processed data model to learn from as K-NN, SVM and ANN score 27.19%, 36.87% and 32.72%, respectively.

Table 2. Accuracies of the models using SMOTEBoost as pre-processing.

| Models      | Accuracy (%) | Precision (%) | Recall (%) | F measure (%) |
|-------------|--------------|---------------|------------|---------------|
| ANN + PCA   | 70.05        | 55.96         | 59.21      | 57.54         |
| ANN + SOM   | 34.56        | 28.55         | 25.22      | 26.78         |
| ANN + GHA   | 70.51        | 63.08         | 48.37      | 54.75         |
| ANN         | 71.43        | 64.43         | 51.32      | 57.13         |
| K-NN + PCA  | 35.35        | 17.02         | 22.25      | 19.29         |
| K-NN + SOM  | 30.41        | 30.84         | 30.89      | 30.86         |
| K-NN + GHA  | 15.67        | 13.39         | 15.89      | 14.53         |
| K-NN        | 47.00        | 45.45         | 47.24      | 46.33         |
| SVM + PCA   | 76.50        | 69.22         | 56.02      | 61.92         |
| SVM + SOM   | 34.10        | 22.26         | 24.74      | 23.43         |
| SVM + GHA   | 74.65        | 70.39         | 53.83      | 61.01         |
| SVM         | 69.59        | 66.42         | 48.63      | 54.93         |

Table 3. Performance of the models using SMOTEBoost and OWB as pre-processing.

| Models      | Accuracy | Precision | Recall | F measure |
|-------------|----------|-----------|--------|-----------|
| ANN + PCA   | 77.01    | 81.77     | 87.50  | 84.54     |
| ANN + SOM   | 56.57    | 56.19     | 61.13  | 58.56     |
| ANN + GHA   | 69.87    | 69.48     | 75.55  | 72.39     |
| ANN         | 84.97    | 89.84     | 85.17  | 87.44     |
| K-NN + PCA  | 63.74    | 63.92     | 54.77  | 58.99     |
| K-NN + SOM  | 68.49    | 69.61     | 74.06  | 71.77     |
| K-NN + GHA  | 57.02    | 89.82     | 63.09  | 74.12     |
| K-NN        | 65.83    | 72.73     | 65.17  | 68.74     |
| SVM + PCA   | 76.17    | 86.85     | 62.75  | 72.86     |
| SVM + SOM   | 39.77    | 41.36     | 45.23  | 43.21     |
| SVM + GHA   | 75.71    | 81.21     | 66.51  | 73.13     |
| SVM         | 78.48    | 84.32     | 64.80  | 73.28     |

SMOTEBoost

As soon as SMOTEBoost is applied, accuracy for ANN increases by 1% for PCA and for GHA it boosts up by almost 8% as shown in Table 2. K-NN is again the worst performer here; the only significant accuracy boost comes from using SOM which is about 3%. For every other model K-NN has lower accuracy. SVM also has a reduced accuracy for PCA and no dimensionality reduction; reduction of about 1% and 7%, respectively. However, SVM has gained about a 10% higher accuracy for the data model pre-processed using GHA.
**SMOTEBoost and optimal width binning**

As soon as optimal width binning is applied, the problem of overfitting in $K$-NN perishes and this reflects in its result as it achieves 68.49%, 65.83% and 63.74% for SOM, no dimensionality reduction and PCA, respectively, as in Table 3, a much higher accuracy compared to its previous models. SVM performs well and produces reliable results with high accuracies of 78.48% for no dimensionality reduction and about 76% for both PCA and GHA. ANN is the best performer for this model, achieving an accuracy of about 85% for no dimensionality reduction with very high precision, recall and $F$ measure; this means the quality of its prediction is higher.

**SMOTEBoost with optimal width binning and normalization**

For this data model, $K$-NN has managed to just reach the 70% accuracy mark to produce a considerably better model than the ones it generated with lower pre-processed data. SVM performs well again, achieving 75.93% and 75.71% for PCA and no dimensionality reduction respectively as shown in Table 4. Here, it is observed again that ANN is the best performer with an improvement of over 2% from the previous model to achieve the highest accuracy so far; a highly commendable 86.11%. It can be inferred that for SOM, the methodology to reduce the data into a mere two dimensions destroys the integrity and quality of the data set and in turn it does not favour any of the learners. The same could be said about GHA since it uses unsupervised learning in a similar way to SOM but it performs slightly better due to its preference to work more like PCA. We will discard SOM and GHA because PCA has significantly performed better to predict HSC result. Thus, we will use the models with PCA and without any dimensionality reduction for the remainder of our result analysis.

**Deriving the best models**

We have derived the best models through various pre-processing techniques. We also find the best pre-processing technique for each of the models. A comparative study shown in Tables 5 and 6. After performing all the comparisons, we incorporated the best results of all the models in Table 7.

### Table 4. Performance of the models using SMOTEBoost with OWB and normalization.

| Models       | Accuracy (%) | Precision (%) | Recall (%) | $F$ measure (%) |
|--------------|--------------|---------------|------------|-----------------|
| ANN + PCA    | 77.78        | 75.68         | 68.95      | 72.16           |
| ANN + SOM    | 52.90        | 53.91         | 57.15      | 55.48           |
| ANN + GHA    | 72.62        | 72.46         | 78.54      | 75.38           |
| ANN          | 86.11        | 89.72         | 85.18      | 87.39           |
| $K$-NN + PCA | 70.00        | 82.37         | 61.48      | 70.41           |
| $K$-NN + SOM | 62.99        | 71.52         | 53.63      | 61.30           |
| $K$-NN + GHA | 62.59        | 62.69         | 67.60      | 65.05           |
| $K$-NN       | 69.31        | 71.30         | 63.74      | 67.31           |
| SVM + PCA    | 75.93        | 85.43         | 57.36      | 68.64           |
| SVM + SOM    | 33.78        | 33.00         | 38.38      | 35.49           |
| SVM + GHA    | 66.96        | 79.24         | 62.99      | 70.19           |
| SVM          | 75.71        | 85.48         | 64.94      | 73.81           |
Artificial neural network

ANN achieves accuracies of over 70% for each data model when PCA is applied as shown in Table 5; with no pre-processing, we have got 70.05% accuracy and after applying SMOTEBoost technique to balance our data set along with OWB it boosts its accuracy to 77.01%. It achieves a highest of 77.78% after pre-processing using SMOTEBoost + OWB + Normalization. From Table 6 it can be observed that, without any dimensionality
reduction, ANN performs the best, achieving accuracies of 70.05% without any pre-processing technique. As we have applied SMOTEBoost + OWB, the accuracy burgeons to 84.97%. Lastly, the accuracy augments to 86.11% as normalization is introduced along with SMOTEBoost + OWB.

**K-nearest neighbour**

Table 5 shows that K-NN is the least reliable in generating the predictive model when used with PCA; however, from the first data model to the last, K-NN has seen an increase in accuracy using PCA; its accuracy started with about 30%, then 35% after SMOTEBoost, 63% after SMOTEBoost + OWB and after adding Normalization to the mix, it soared to 70%. This rapid increase in accuracy and overcoming the problem of overfitting for K-NN shows that our techniques in pre-processing have been nothing short of exemplary. K-NN performs unimpressively, scoring just 65.83% and 69.31% when used with SMOTEBoost + OWB only and when used with SMOTEBoost + OWB + Normalization respectively for no dimensionality reduction as shown in Table 6.

**Support vector machine**

In Table 5, it can be observed that SVM achieves accuracies of over 70% for each data model when it is used in junction with PCA, achieving a highest accuracy of 77.88% for no pre-processing. Notwithstanding achieving highest accuracy without any pre-processing its F measure has notably augmented from 62.69% to 72.86% as we have introduced different pre-processing techniques. Increasing the F measure indicates that the quality of the predictive accuracy has significantly increased. It achieves a high accuracy, topping off at 78.48% when used with SMOTEBoost + OWB when no dimensionality reduction is used as shown in Table 6 (Figure 4).

SVM is the second best learner to generate our predictive model as our study suggests; however, considering the target class is continuous, accuracies above 70% are rather excellent. SVM does perform better with PCA, as shown in Table 7, except for the data models which are pre-processed by SMOTEBoost + OWB + Normalization and SMOTEBoost + OWB only; the differences being 0.22% and 2.91%, respectively. In the other data models SVM shows improvement in accuracies when using PCA vs. no dimensionality reduction.

![Figure 4](image-url). The effect of pre-processing on each of the learners when keeping PCA as the constant dimensionality reduction technique.
Conclusion

From these final models in Tables 5 and 6 and a breakdown of dimensionality reduction comparison in Table 7, it can be observed that ANN performs the best with an accuracy of over 86%. In most cases dimensionality reduction plays a part in reducing the integrity of the data and in turn reduces accuracies for most models except for some cases such as SVM without any pre-processing and SVM with SMOTEBoost when used with PCA; from this it can be inferred that due to PCA generating orthogonal attributes, the learner performs better.

K-NN shows how data pre-processing can help improve the quality of the data; it started off with about 30% accuracy and moved up to 70% as more data pre-processing was applied. While it may not be the best fit to produce a predictive model for this particular study, it demonstrates the importance of data pre-processing when building predictive systems. Parents and the responsible authorities can gather knowledge from the patterns of the HSC performance through our study.

One of the major limitations of our work is that there is no way to validate the responses from the survey. Another limitation is that the study only targets predicting the overall HSC grade. These limitations provide groundwork for future research: survey response validation and subject level grade prediction. Our research would augment the possibility of performing better in a course by knowing the possible grade beforehand through elevating the awareness relevant to that particular course(s). Survey response validation is trickier to solve but if a data set’s integrity could be proved then it would provide reliable predictions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

A. T. M. Shakil Ahamed is working as a part-time Lab Officer at the Electrical and Computer Engineering Department, North South University, Dhaka, Bangladesh. He received his BSc in
Computer Science and Engineering from North South University, Dhaka, Bangladesh in 2015. He has 4 conference proceedings in the areas of fuzzy expert system, data mining and machine learning.

Navid Tanzeem Mahmood is working as a part-time Lab Officer at the Electrical and Computer Engineering Department, North South University, Dhaka, Bangladesh. He received his BSc in Computer Science and Engineering from North South University, Dhaka, Bangladesh in 2015. He has 4 conference proceedings in the areas of fuzzy expert system, data mining and machine learning.

Rashedur M. Rahman is working as a Professor at the Electrical and Computer Engineering Department, North South University, Dhaka, Bangladesh. He received his PhD in Computer Science from the University of Calgary, Calgary, AB, Canada, and MS degree from the University of Manitoba, Winnipeg, MB, Canada, in 2007 and 2003, respectively. He has authored more than 130 peer-reviewed journals and conference proceedings in the area of parallel, distributed, grid, and cloud computing, knowledge, and data engineering. He has been serving in the editorial board of a number of journals in the knowledge and data engineering field. He has been serving as the Computer Science Area Editor from Bangladesh for Hindawi Publishing Corporation. He has been also serving as an organizing committee member for various international conferences organized by IEEE, ACM, Springer. He also served as a reviewer of a couple of journals published by Elsevier, Springer and Wiley. He is currently a Fellow and formal Counsellor of Bangladesh Computer Society.

References

Aatta, M. A., & Jamil, A. (2012). Effects of motivation and parental influence on the educational attainments of students at secondary level. *Academic Research International*, 2(3), 427–431.

Adeyemo, A. R., Oladipipo, A., & Omisore, A. O. (2013). Teachers’ motivation on students’ performance in mathematics in government secondary schools, Makurdi Lg area. *International Journal of Humanities and Social Science Invention*, 2(5), 35–41.

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist*, 49(1), 15–24.

Aertgeerts, B., & Buntinx, F. (2002). The relation between alcohol abuse or dependence and academic performance in first-year college students. *Journal of Adolescent Health*, 31(3), 223–225. doi:10.1016/s1054-139x(02)00362-2

Ahamed, A., Mahmood, N. T., & Rahman, R. M. (2016). Prediction of HSC examination performance using socioeconomic, psychological and academic factors. In 9th International Conference on Electrical and Computer Engineering (ICECE), IEEE. doi:10.1109/ICECE.2016.7853906

Ahamed, A., Mahmood, N. T., & Rahman, R. M. (2017). Prediction of academic performance during adolescence based on socioeconomic, psychological and academic factors. In D. Król, N. T. Nguyen, & K. Shirai (Eds.), *Advanced topics in intelligent information and database systems* (pp. 71–80). *Studies in computational intelligence*. Kanazawa: Springer. doi:10.1007/978-3-319-56660-3_7

Alkhatib, K., Najadat, H., Hemeidi, I., & Ali Shatnawl, M. K. (2013). Stock price prediction using K-nearest neighbor (K-NN) algorithm. *International Journal of Business, Humanities and Technology*, 3(3), 32–44.

Anderson, G., Benjamin, D., & Fuss, M. A. (1994). The determinants of success in university introductory economics courses. *The Journal of Economic Education*, 25(2), 99–119. doi:10.2307/1183277

Aripin, R., Mahmood, Z., Rohaizad, R., Yeop, U., & Anuar, M. (2008). *Students’ learning styles and academic performance 22nd annual SAS Malaysia forum*. Kuala Lumpur: Kuala Lumpur Convention Center.

Atta, M. A., Jamil, A., Rehman, J., Ayaz, M., Saeed, A., & Shah, M. A. (2011). Effects of private home tuition on educational attainments of students at secondary school level. *World Applied Sciences Journal*, 13(6), 1486–1491.

Ayers, E., Nugent, R., & Dean, N. (2009). *A comparison of student skill knowledge estimates* (pp. 1–10). International conference on Educational Data Mining, Cordoba.
Behrman, J. R., & Lavy, V. (1998). Child health and schooling achievement: Association, causality and household allocations (CARESS Working Papers, 97–23).

Brunstein, J. C., & Gollwitzer, P. M. (1996). Effects of failure on subsequent performance: The importance of self-defining goals. *Journal of Personality and Social Psychology*, 70(2), 395–407. doi:10.1037/0022-3514.70.2.395

Buckles, K. S., & Munnich, E. L. (2012). Birth spacing and sibling outcomes. *Journal of Human Resources*, 47, 613–642. doi:10.3368/jhr.47.3.613

Chawla, N. V., Lazarevic, A., Hall, L. O., & Bowyer, K. W. (2003). SMOTEBoost: Improving prediction of the minority class in boosting. *Knowledge Discovery in Databases: PKDD*, 107–119. doi:10.1007/978-3-540-39804-2_12

Cheers, B. (1990). Rural disadvantage in Australia. *Australian Social Work*, 43(1), 5–13.

Christian, K., Morrison, F. J., & Bryant, F. B. (1998). Predicting kindergarten academic skills: Interactions among child care, maternal education, and family literacy environments. *Early Childhood Research Quarterly*, 13(3), 501–521. doi:10.1016/s0885-2006(99)80054-4

Coleman, J. S., Ernest, Q., Campbell, C. J., Hobson, J. M., Alexander, M. M., Frederick, D. W., & Robert, L. Y. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Department of Health Education and Welfare. Office of Education. *Computers in Education*, Washington, DC (pp. 1–5).

Ellore, S. B., Niranjan, S., & Brown, III, U. J. (2014). The influence of Internet usage on academic performance and face-to-face communication. *Journal of Psychology and Behavioral Science*, 2(2), 163–186.

Flad, K. (2010). *The influence of social networking participation on student academic performance across gender lines* (Master’s thesis). The College at Brockport. Retrieved from http://digitalcommons.brockport.edu/cgi/viewcontent.cgi?article=1030&context=edc_theses

Garrett, B. E., Dube, S. R., Trosclair, A., Caraballo, R. S., & Pechacek, T. F. (2011). Cigarette smoking – United States, 1965–2008, National Center for Chronic Disease Prevention and Health Promotion. *CDC*, 60(01), 109–113.

Gedeon, T. D., & Turner, H. S. (1993). *Explaining student grades predicted by a neural network* (pp. 609–612). International conference on Neural Networks, Nagoya.

Gorrel, G. (2006). *Generalized Hebbian algorithm for incremental singular value decomposition in natural language processing*. Retrieved from http://www.aclweb.org/anthology/E06-1013

Halpern, C. T., Joyner, K., Udry, J., & Suchindran, C. (2000). Smart teens don’t have sex (or kiss much either). *Journal of Adolescent Health*, 26(3), 213–225. doi:10.1016/s1054-139x(99)00061-0

Hatzichristou, C. (1993). Children’s adjustment after parental separation: Teacher, peer and self-report in a Greek sample: A research note. *Journal of Child Psychology and Psychiatry*, 34(8), 1469–1478. doi:10.1111/j.1469-7610.1993.tb02104.x

Islam, M. N., & Pavel, T. (2011). Factors contributing towards dropouts at undergraduate level: An analysis. *ASA University Review*, 5(1), 97–113.

Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, 3(1), 89–93. doi:10.7763/ijcte.2011.v3.288

Jeynes, W. H. (2002). Examining the effects of parental absence on the academic achievement of adolescents: The challenge of controlling for family income. *Journal of Family and Economic Issues*, 23(2), 189–210.

Jimenez, L. O., & Landgrebe, D. A. (1998). Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 28(1), 39–54.

Jishan, S. T., Rashu, R. I., Mahmood, A., Billah, F., & Rahman, R. M. (2015). Application of optimum binning technique in data mining approaches to predict students’ final grade in a course. *Advances in Intelligent Systems and Computing*, 331, 159–170. doi:10.1007/978-3-319-13153-5_16

Johnson, R. B. (1996). *The effectiveness of preschool education on academic achievement* (p. 143). Retrieved from U.S. Department of Education, ERIC website: http://files.eric.ed.gov/fulltext/ED400069.pdf

Jolliffe, I. T. (2002). *Principal component analysis*. Springer series in statistics, (2nd ed.). New York, NY: Springer.
Kar, A. (n.d.). Stock prediction using artificial neural networks. Dept. of Computer Science and Engineering, IIT, Kanpur. Retrieved from http://people.eecs.berkeley.edu/~akar/IITK_website/EE671/report_stock.pdf.

Kayah, F. (2008). Discretizing continuous features for naive Bayes and C4. 5 classifiers. College Park: University of Maryland Publications.

Keeley, T. J., & Fox, K. R. (2009). The impact of physical activity and fitness on academic achievement and cognitive performance in children. International Review of Sport and Exercise Psychology, 2(2), 198–214. doi:10.1080/17509840903233822

Kirby, A., & McElroy, B. (2003). The effect of attendance on grade for first year economics students in University College Cork. The Economic and Social Review, 34(3), 311–326.

Kirk, A. M. (2000). Riding the bull: Experience with individual market reform in Washington, Kentucky, and Massachusettts. Journal of Health Politics, Policy and Law, 25(1), 133–173.

Kohonen, T., & Honkela, T. (2007). Kohonen network. Scholarpedia, 2(1), 1568. doi:10.4249/scholarpedia.1568

Massoni, E. (2011). Positive effects of extra-curricular activities on students. ESSAI, 9, article 27.

McFadden, K., & Dart, J. (1992). Time management skills of undergraduate business students. Journal of Education for Business, 68(2), 84–88. doi:10.1080/08832323.1992.10117592

Min, J., & Lee, Y. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications, 28(4), 603–614. doi:10.1016/j.eswa.2004.12.008

Mullin, M., & Sukthankar, R. (2002). Complete cross-validation for nearest neighbor classifiers. Retrieved from http://www.cs.cmu.edu/~rahuls/pub/icml2000-rahuls.pdf

Odok, A. O. (2013). Contemporary family structures and students’ academic performance in secondary schools in Ikom Local Government Area, Cross River State. Journal of Sociology, 4(4), 87–94.

Oladokun, V. O., Adebunjo, A. T., & Charles-owaba, O. E. (2008). Predicting student’s academic performance using artificial neural network: A case study of an engineering course. Pacific Journal of Science and Technology, 9(1), 72–79.

Pal, A. K., & Pal, S. (2013). Analysis and mining of educational data for predicting the performance of students. International Journal of Electronics Communication and Computer Engineering, 4(5), 1560–1565.

Schoon, I., & Parsons, S. (2002). Teenage aspirations for future careers and occupational outcomes. Journal of Vocational Behavior, 60(2), 262–288.

Usaini, M. I., & Abubakar, N. B. (2015). The impact of parents’ occupation on academic performance of secondary school students in Kuala Terengganu. Multilingual Academic Journal of Education and Social Sciences, 3(1), 115.

Wang, T., & Mitrovic, A. (2002). Using neural networks to predict student’s performance. IEEE, 112–120. doi:10.1109/CIE.2002.1186127

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Data mining applications: A comparative study for predicting student’s performance. arXiv preprint arXiv:1202.4815.