An artificial neural network approach in predicting career strand of incoming senior high school students

A L Nazareno¹, M J F Lopez², G A Gestiada¹, M P Martinez¹ and R M Roxas-Villanueva¹

¹Institute of Mathematical Sciences and Physics, College of Arts of Sciences, University of the Philippines Los Baños
²Institute of Statistics, College of Arts of Sciences, University of the Philippines Los Baños

Email: alnazareno@up.edu.ph

Abstract. The K to 12 program has been implemented in the Philippines by the Department of Education which implicated an additional two years in the students’ basic education. These ancillary years allow senior high school students to take courses under the core curriculum and the track of choice. Each student must select one track to pursue that can equip him/her with skills to prepare for the future. Prediction of choice of career track in senior high school is advantageous for educational institutions since it gives insights that can help them develop vital programs beneficial for students’ learning in school. In this study, we applied artificial neural network (ANN) to predict the career strand based on the students’ grades in five major subjects. Different ANN models have been considered and compared. In training and testing the models, a sample of 293 student data information was used. The highest accuracy recorded among all the models was 74.1%.

1. Introduction

The K to 12 program administered by the Department of Education (DepEd) was implemented in the Philippines with an addition of two years for senior high school (SHS). The SHS includes four different tracks – Academic, Technical-Vocational-Livelihood (TVL), Sports, and Arts and Design. Under the academic track, there are four strands – General Academic Strand (GAS), Science, Technology, Engineering, and Mathematics (STEM), Accountancy, Business, and Management (ABM), and Humanities and Social Sciences (HUMSS). For the TVL track, there are Agri-fishery Arts, Home Economics, Industrial Arts, and Information and Communication Technology (ICT) as strands [1].

Since the program is a relatively new initiative of the Philippine government, it is beneficial and strategic to conduct research studies on K to 12. Currently, there are only a few studies published concerning K to 12 in the Philippines. Some studies include the works of [2-5].

Determining the career track that a student may want to pursue in SHS is favorable to the school administrators, teachers, and students. With this information, it is possible to project future enrollees in a particular track. This program can be of significant help to DepEd and other private education institutions in managing resources and in executing their plans and policies. The study can also help in identifying important factors relating to career track decision which can guide teachers in developing
instructional materials and pedagogical strategies that might truly help students perform well in their chosen track.

Various studies determined the correlation between academic performance of students to their chosen career path. The findings of the study of Ahmed [6] revealed that students with high levels of anxiety in Mathematics are more likely to choose career unrelated to STEM strand. A study [7] investigated the motivations of academically excellent students in choosing careers related to teaching. It was found out that there is a positive reciprocal association between career planning and academic achievement [8]. Furthermore, the Brainard’s Occupational Preference Inventory (BOPI) results are significantly related to the elective course grades of students in University of Rizal System’ Laboratory School [9].

In this study, we constructed a model based on Artificial Neural Network (ANN) that can predict the chosen strand of students using their grades in the five subjects. ANN is biologically-inspired programming framework that allows the processing of information (i.e. inputs) for extraction of patterns and detection of trends. Particularly, a multilayer perceptron (MLP), a class of feedforward ANN that utilizes a supervised learning approach, was used to predict the chosen strand. Furthermore, the grades on various academic subjects were considered as input in the ANN model. A detailed discussion of the ANN framework is given in the succeeding section.

2. Artificial Neural Network (ANN)
A neural network is an artificial representation of the human brain that tries to simulate its learning process. One characteristic of ANN that makes it very useful is its ability to perform tasks involving complex systems. It is apparent that not all relationships existing in a pool of data are linear in nature. It is notably useful in educational context since learning and other related concepts in education [10-13], like career decision making, are complex processes.

The ANN model is similar to the structure of the nervous system and it follows how biological cognition works. The neuron is the basic unit of the nervous system which serves as a processing element. A neuron is composed of three core components such as dendrites, cell body or soma, and axon. Dendrites receive signals from external environment or from the outputs of other neurons. The signals are then sent to the soma which are transferred to the axon and consequently conveys to other dendrites of the adjacent neurons via the synapses. An aggregation of all these neurons and the corresponding neural processes forms the neural network [14,15].

Artificial neural networks can be trained to predict the output (e.g. career strand) given certain inputs (e.g. grades on the different subjects). A multilayer perceptron (MLP) is a class of feedforward artificial neural network. Multi-layer perceptron (MLP) is a multi-layer feed forward network trained by supervised technique called the backpropagation algorithm. It constructs the model based on examples of data. The detailed process of the backpropagation algorithm for a multilayer perceptron with one hidden layer is described in [16].

2.1. Discretization
Machine learning methods often prefer relatively small number of attribute values [17]. Sample space reduction is important to accompany the capability of the software and hardware to handle data analysis of large data set [18]. It is necessary to transform input data from continuous to discrete especially when you are dealing with large data set with complex [17, 20] but with minimal loss of information [20]. When most of the features are numeric, it may be necessary to make the data discretized which can also be used when considering decision system [21].

2.2. Correlation Analysis
Selecting correlated features is essential in finding a good machine learning model. Several methods can be used to determine the relationship between two variables. The method depends on the type/level of measurement of the variables and the distribution of the variable(s) being examined. In particular, for quantitative variables, the joint distribution of the two variables should be bivariate normal. If this assumption is not satisfied, a non-parametric test (distribution-free test) can be used as an alternative.
The following are used to determine the strength and direction of relationship between two variables based on their type and level of measurement:

a. Pearson’s Product Moment Correlation Coefficient
The Pearson’s Product Moment Correlation Coefficient is used to determine the strength and direction of linear relationship between two quantitative variables, say X and Y, that are at least in the interval scale. It assumes that the two variables are bivariate normal.

b. Spearman Rank Correlation Coefficient
The Spearman Rank Correlation Coefficient can be used when the assumption on normality is not satisfied and can be used to test the relationship between two variables that are at least in the ordinal level of measurement.

c. Point Biserial Correlation Coefficient
The Point Biserial Correlation Coefficient is a special case of the Pearson’s Correlation Coefficient. It is being used to measure the degree and direction of relationship between a continuous variable (at least interval in scale) and a nominal variable which only takes two values.

d. Chi-square based measures
Chi-square based measures are used to determine the strength of relationship between two qualitative variables. This is being obtained when the Chi-square test of independence yields to a significant result which indicates association between the two variables being tested.

3. Methodology

3.1. Data collection
Data from 293 grade 11 students from STEM (N1=157) and GAS (N2=136) were considered. The respondents are composed of 144 males and 149 females with an average age of about 17 years old. Demographic information of the students and their grades obtained in various subjects during their Grade 10 year were obtained. The subjects include Filipino, English, Mathematics, Science, and Technology and Livelihood Education (TLE).

3.2. Discretization of data set
The original data set was transformed by categorizing the numeric grades as low, average, and high. The grade bins were determined by considering mean grade plus/minus one-half standard deviation. The binning process used is considered as a norm-referenced interpretation of performance since the students were categorized by comparing their numeric grade to the mean numeric grade of the population in each subject [22]. Norm-referenced measure is normally used for selection decisions i.e. categorizing student performance to low, average and high, since it distinguishes between performances of students through a range of scores [23].

3.3. Correlation analysis
Correlation analysis was conducted to both the original data set and the discretized data set. In particular, the point-biserial correlation analysis was used to determine the strength of association between the grades and the chosen career strand following the Pearson’s product moment interpretation of the correlation coefficient.

On the other hand, the Chi – square test determined the association between the discretized grades and the chosen career strand. The computed Cramer’s V coefficient was used to identify the strength of association between the two variables. Coefficients less than 0.1 indicate weak association, while values greater than 0.1 but less than 0.3 correspond to a moderate association. Further, coefficients greater than 0.3 signify strong association.
3.4. Construction of the ANN predictive model

The basic architecture of ANN consists of 3 layers: input layer, hidden layers, and output layer, as shown in Figure 1. Each layer is made up of neurons that are connected to the neurons of the adjacent layer.

In this study, the ANN network architecture considered was composed of 1 input layer, 1 or 2 hidden layers, and 1 output layer. There are 5 neurons for the input layer pertaining to the grades of the student in five subjects specified and one neuron for the output layer indicating the type of strand. If the students’ career strand is STEM, the value assigned is 1. On the other hand, if the students’ career strand is GAS, the value assigned is 0. The number of hidden neurons for the hidden layers (either 1 or 2) was determined during the implementation of ANN (see the next step).

3.5. Implementation of the ANN algorithm

A python-based program was developed to implement the ANN algorithm. Various packages of python such as pandas, numpy, and sklearn among others were utilized. The program included the following sections:

a. Loading of the data: The data sets were loaded in the program.

b. Assignment of input and output data: After loading the data, input and output data were assigned appropriately.

c. Normalization of the data: Normalization scheme was used primarily to make all the input/output values at a comparable range that can help improve the implementation of the algorithm.

d. Fine-tuning of hyperparameters: Fine-tuning was used to find the best combination of parameters of the ANN model. Table 1 summarizes the parameters for the ANN model.

![Figure 1. The basic architecture of ANN.](image)

| Hyperparameter                     | Range of values     | Suggested Readings |
|------------------------------------|---------------------|--------------------|
| Activation Function for the Hidden Layers | {identity, logistic, tanh, relu} | [24, 25]          |
| Solver for Weight Optimization     | {lbfgs, sgd, adam}  | [26-28]            |
| Learning Rate Schedule for Weight Update | {constant, invscaling, adaptive} | [25]            |
| No. of Hidden Layers               | {1,2}               | [25, 29]           |
| No. of Neurons/Nodes in the Hidden Layer | {1,2, 3,...,20}   | [25, 29]           |
Each dataset was divided into training and testing sets using the train_test_split utility function. GridSearchCV module was employed to generate various combinations of hyperparameters such as number of hidden layers, type of activation function, solver, learning rate, and number of neurons in the hidden layer/s. The mean accuracy rates were calculated from the 10-fold cross-validation procedures and compared to determine the best set of hyperparameters. The hyperparameters obtained were used to establish the predictive model.

4. Results and discussion

4.1. Establishing the relationship between the grades and the career strand
Correlation analysis showed the direction and strength of association between the grades and the chosen career strand. As indicated in Table 2, the computed correlation coefficients for all the subjects suggest a positive weak relationship between grades and the chosen career strand. The observed relationship is further shown in the scatter plots generated, which revealed positive slopes implying that students with high grades are more likely to be associated to students who chose the STEM strand and vice versa. The results also showed that Filipino, English and Math are among the subjects which are perceived to have high correlation on the choice of career strand accounting about 14.75%, 12.53%, and 10.69% in the total variation on the choice of career strand, respectively. However, for predictive purposes, this contribution might not be of big influence.

Because of the low contribution to the total variation in the choice of career strand of the original features, the grades were transformed to grade bins by obtaining the mean grade plus/minus one-half standard deviation. Histograms shown in Figure 2 were also used as basis for the transformation used. It can be observed that all grades exhibited similar distribution and performance of the students can be classified as low, average and high. Hence, grades within the mean plus/minus the standard deviation are considered as average while grades lying outside the mean minus the standard deviation and mean plus the standard deviation are considered low and high, respectively. Correlation analysis using the Chi-Square Test of Independence were performed on the discretized values and the results are shown in

![Figure 2. Distribution of the student’s grade in the five subjects.](image-url)
The results of the analysis showed that the discretized grades have moderate to strong association to the chosen career strand. Further scrutiny of the contingency tables, indicates that students who get high grades in most of the subjects, tend to choose the STEM strand. The relationship among the grades and the strand are further illustrated in the biplot in Figure 3 which shows that students with low to average grades are associated to GAS while students with average to high grades are likely to be associated to STEM strand. Among the five subjects Filipino, English and Math can be considered as important features in classification of the chosen career strand because of their high computed Chi-square statistic.

### Table 2. Relationship of the student’s grade and his/her chosen career strand.

| Variable | Rho   | R     | Sig     |
|----------|-------|-------|---------|
| Filipino | 0.384 | 14.75 | <0.0001 |
| English  | 0.354 | 12.53 | <0.0001 |
| Math     | 0.327 | 10.69 | <0.0001 |
| Science  | 0.224 | 5.02  | <0.0001 |
| TLE      | 0.258 | 6.66  | <0.0001 |

### Table 3. Relationship of the discretized student’s grade and chosen career strand.

| Variable | Chi-Square | Sig  | Cramer’s V  |
|----------|------------|------|-------------|
| Filipino | 39.8036    | 0.000 | 0.3686      |
| English  | 48.6650    | 0.000 | 0.4075      |
| Math     | 30.4112    | 0.000 | 0.3222      |
| Science  | 17.8632    | 0.000 | 0.2469      |
| TLE      | 16.8587    | 0.000 | 0.2399      |

### Figure 3. Biplot of the grades of students in the five subjects and the chosen career strand.

4.2. **ANN implementation and performance assessment**

The best combinations of the hyperparameters were identified by comparing the mean accuracy rates from the 10-fold cross-validation procedure. Two combinations for each dataset were generated, for one hidden layer and for two hidden layers, resulting to 4 models.

As highlighted in Table 4, the best model is the architecture with two hidden layers where the activation function for the hidden layers is logistic sigmoid function. According DasGupta & Schnitger [30], this function is useful especially in networks trained by the back-propagation algorithm for it is easily differentiable and in return, can minimize the computation capacity for training.
Table 4. Results of the 10-fold cross-validation procedure.

| No. of Hidden Layer/s | Data Set | Activation Function | Solver | Learning Rate | No. of Nodes for Each Layer | Accuracy Rate |
|-----------------------|----------|---------------------|--------|---------------|-----------------------------|---------------|
| One                   | 1        | logistic             | lbfgs  | invscaling    | 6                           | 0.635         |
|                       | 2        | tanh                | adam   | constant      | 8                           | 0.726         |
| Two                   | 1        | tanh                | lbfgs  | invscaling    | (8,19)                      | 0.630         |
|                       | 2        | logistic             | lbfgs  | constant      | (7,2)                       | 0.740         |

In training the ANN, Limited-memory Broyden-Fletcher-Goldfarb-Shanno, solver for weight optimization, and constant learning rate were utilized to obtain the highest mean accuracy rate. According to Liu & Nocedal [26], Limited-memory Broyden-Fletcher-Goldfarb-Shanno is very competitive due to its low iteration cost. Though keeping the learning rate constant does not guarantee meaningful convergence [31] since the default initial learning is 0.001, learning rate annealing may not be necessary, which could explain why the learning rate schedule is set constant all throughout the training process.

The model with two hidden layers yielded the best accuracy with 7 hidden neurons in the first and 2 hidden neurons in the second layer. A summary of the methods on finding the optimal hidden neurons was given in [32], however for this study, we just adopted the results obtained from the 10-fold cross-validation.

The best model was trained fifty times generating mean accuracy rates of 74.1 % and 68.6 %, for the training set and the testing set, respectively. This is better than the performance of the models which utilized the original dataset.

A ROC curve was generated for best ANN model. As shown in Figure 4, the ROC curve lies above the 45-degree line which indicates that percentage of the model correctly predicting the chosen career strand is higher than the percentage of incorrect prediction. This means that the model (with grades as parameter) can discriminate well those students who chose the STEM strand and those who chose the GAS. On the average, the model can give 79% correct prediction among all STEM-GAS pairs in the dataset.

![Figure 4. ROC curve of the best ANN model.](image)

Based on the results shown in Table 4, the performance of ANN models (for 1 or 2 hidden layers) utilizing the transformed data set significantly increased. This improvement can be attributed to the binning scheme. As mentioned in section 3.2, binning method has shown to improve the performance of machine learning algorithms since these methods typically prefer a low number of attribute values.

When the features of the original data were transformed, the strength of correlation notably increased. Correlation analysis is one way to filter data and select important features. Moderate to strong
correlations indicate a good choice of features. Training with features with very little correlation may yield to inaccurate results, which happened when the original dataset was used.

We further examined the transformed data set by investigating the distribution of the features. The generated boxplots show that the grades in TLE has an outlier; presence of outliers can possibly distort the training process and lead to a less accurate result. The transformed data was analyzed and revealed the removal of these outliers.

5. Summary and recommendation
We were able to develop a predictive model that can determine the choice of strand of an incoming senior high school student with an accuracy of 74.1 %. This study was able to show the use of discretization method to increase the accuracy of the prediction. Grades in the five core courses were incorporated as inputs in training the ANN model.

The study is a significant initiative to utilize the AI-driven method in predicting the possible career strand choice of an incoming Grade 11 student. This can be extended by considering other factors such as income of the family, parents’ and siblings’ educational background, among others. Moreover, the other tracks or strands can be incorporated as well in the analysis. Other AI-driven techniques such as support vector machines, k nearest-neighborhood, logistic regression, random forest, or combination of these methods can be employed to develop the predictive model. Another essential offshoot of this study is the possible creation of an automated predictive system that can be provided to the schools that can help them in administering academic policies.

Acknowledgement
This work was funded by the UP System Enhanced Creative Work and Research Grant (ECWRG 2017-2-009).

References
[1] Anon K to 12 Basic Education Curriculum Republic of the Philippines: Department of Education
[2] Cabansag M 2014 Impact statements on the k-12 science program in the enhanced basic education curriculum in provincial schools Researchers World 5 29–39
[3] Gestia G, Nazareno A and Roxas-Villanueva R M 2017 Development of a senior high school career decision tool based on social cognitive career theory Philipp. J. Sci. 146 445–55
[4] Mohammad N K 2016 The perception of the parents and students on the implementation of k – 12 basic education program in the Philippines, Int. Conf. on Education (IECO) Proc. 1 481-503
[5] Okabe M 2013 Where does Philippine education go? the k to 12 program and reform of Philippine basic education Institute of Developing Economies 425 1–30
[6] Ahmed W 2018 Developmental trajectories of math anxiety during adolescence: associations with STEM career choice J. Adolesc. 67 158–66
[7] Kass E and Miller E C 2018 Career choice among academically excellent students: choosing teaching career as a corrective experience Teaching and Teacher Education 73 90–8
[8] Negru-Subirica O and Pop E I 2016 Longitudinal links between career adaptability and academic achievement in adolescence J. Vocat. Behav. 93 163–70
[9] Pascual N T 2014 Factors affecting high school students’ career preference: a basis for career planning program Int. J. Sci.: Basic Appl. Res. 16 1–14
[10] Dharmasaroja P and Kingkaew N 2016 Application of artificial neural networks for predictionof learning performances 12th Int. Conf. on Natural Computation, Fuzzy Systems and Knowledge Discovery 745–51
[11] Karamouzis S T and Vrettos A 2008 An artificial neural network for predicting student graduation outcomes Proc. of the World Congress on Engineering and Computer Science 22–5
[12] Monterola C, Roxas R and Carreon-Monterola S 2009 Characterizing the effect of seating arrangement on classroom learning using neural networks Complexity, 14 26–33
[13] Oancea B, Dragoeescu R and Ciucu S 2013 Predicting students’ results in higher education using neural networks International Conference on Applied Information and Communication Technology 190–3

[14] Jain A K, Mao J and Mohiuddin K M 1996 Artificial neural network: a tutorial Computer 29 31–44

[15] Sánchez Mesa J A, Galán C and Hervás C 2005 The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate Int. J. Biometeorol. 49 355–62

[16] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors Nature 323 533–6

[17] Chmielewski M R and Grzymala-Busse J W 1996 Global discretization of continuous attributes as preprocessing for machine learning Int. J. Approx. Reason. 15 319–31

[18] Yu L, Wang S and Lai K K 2006 An integrated data preparation scheme for neural network data analysis IEEE Trans. Knowl. Data Eng. 18 217–30

[19] Kim K and Han I 2000 Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index Expert Syst. Appl. 19 125–32

[20] Jin R, Breitbart Y and Muoh C 2009 Data discretization unification Knowledge and Information Systems 19 1–29

[21] Borowik G 2013 Boolean function complementation based algorithm for data discretization Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8112 LNCS(PART 2) 218–25

[22] Messick S 1975 The standard problem: meaning and values in measurement and evaluation Am. Psychol. 30 955–66

[23] Mcallister E 1983 Criterion-referenced versus norm-referenced measurement: either or both? Kappa Delta Pi Record 19 58–60

[24] Karlik B and Olgac V 2011 Performance analysis of various activation functions in generalized MLP architectures of neural networks International Journal of Artificial Intelligence And Expert Systems 1 111–22

[25] Basheer I A and Hajmeer M 2000 Artificial neural networks: fundamentals, computing, design, and application J. Microbiol. Methods 43 3–31

[26] Lu D and Nocedal J 1989 On the limited memory BFGS method for large scale optimization Math. Program. 45 503–28

[27] Bottou L 1991 Stochastic gradient learning in neural networks. Proc. Neuro-Nîmes p 91

[28] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization Proc. of 3rd. Int. Conf. for Learning Representations (San Diego) pp 1–15

[29] Zhang G, Patuwo B E and Hu M 1998 Forecasting with artificial neural networks: the state of the art Int. J. Forecast. 14 35–62

[30] DasGupta B and Schnitger G 1992 The power of approximating: a comparison of activation functions Adv. Neural Inf. Process Syst. 5 615–22

[31] Kuan C M and Hornik K 1991 Convergence of learning algorithms with constant learning rates IEEE Trans. Neural Networks 2 484–9

[32] Sheela K G and Deepa S N 2013 Review on methods to fix number of hidden neurons in neural networks Mathematical Problems in Engineering 2013 1-15