

# Predicting Academic Performance with Intelligence, Study Habits and Motivation Factors using Naïve Bayes Algorithm

Kris A. Capao<sup>1</sup>, Anthonette D. Cantara<sup>2</sup>, Angie M. Ceniza<sup>3</sup>,  
Eduardo P. Mendoza Jr. <sup>4</sup>, Stephanie B. Polinar<sup>5</sup>, Joan M. Tero<sup>6</sup>

University of San Carlos  
School of Arts and Sciences  
Department of Computer and Information Sciences

**Abstract** - One of the biggest pedagogical challenges of educators in a computing field is to predict the academic performance of the students. Teachers would like to know which students will pass or fail in order to provide intervention schemes that might help them to pass the subject. The researchers identify some factors to consider in predicting their performance. These factors include their final grades in programming subjects, intelligent quotient, study habits, interpersonal and intrapersonal motivation of students taking computing courses. A semi-supervised machine learning technique called Naive Bayes Algorithm is using in predicting the performance of each students. A data set is divided into training set and testing. The training set is allows the machine to learn a prediction model and the testing set allow the researcher to test the accuracy. With the study, the Naïve Bayes Algorithm accurately predicted a student to pass at 76.18%.

**Keywords:** *Academic Performance, IQ, Study Habits, Motivation, Naïve Bayes, Prediction*

## 1. INTRODUCTION

Academic performance has been studied well by higher education institutions [1]. Many factors are being associated with academic performance and how they correlate and predict the academic performance of a student.

Grades are used as measure of academic performance and the grading system is dependent on the institutions standardized grading which can be numerical or categorical. A summative determination of the academic performance is reflected in the final grade of the students. The basis on continuing the studies is controlled by the final grade. Many latter subjects require pre-requisite subject/s.

Aside from that, in order for the institution to accept a student, the student must complete their admission requirements which include standardized test like Intelligence Quotient (IQ) and others. Being part of the admission requirements can be indicative that it is a factor to the student's academic performance in the institution.

Predicting the academic performance of a student allows the institution to identify the students who are able to graduate on time. This allows the institution to perform intervention. The academic performance of the student reflect to both the student's and the institution's reputation.

## 2. REVIEW OF RELATED LITERATURE

The admission requirements of an institution should be a predictive factor to academic performance. This is the reason of acceptance and disqualification of an applicant to enter the institution. This admission requirement is often paired with other factors like learning style<sup>[2]</sup>, motivation or effort and prior related courses<sup>[3]</sup>. Brazdău and Mihai<sup>[4]</sup> concluded that both consciousness quotient and IQ could better predict the academic performance. Ping<sup>[1]</sup> used the following factors: test anxiety, study habits, math skills and self-efficacy which the author defined as one's belief of having the ability, motivation and resources to complete a task which resulted to self-efficacy and math skills to contribute significantly in the prediction on academic performance. Among the different factors that Verkuyten et. al. <sup>[5]</sup> used; personal motivation became an apparent factor which affects academic performance. Ting<sup>[6]</sup> suggested that with the correlation and manifestation of the other factor in predicting academic success, the institutions may wish to explore other variables for admission decisions.

Wang, et. al.<sup>[7]</sup> used study focus and study duration as one of the factors predicting academic performance. Ahmand, et. al.<sup>[8]</sup>, Affendey et. al.<sup>[9]</sup> and Sundar<sup>[10]</sup> used educational data mining to find patterns and Naïve Bayes as one of the algorithms to predict academic performance.

## 3. METHODOLOGY

The data set is from the students of the University of San Carlos under the Department of Computer and Information Sciences (DCIS). DCIS offers three programs which are Computer Science (BSCS), Information Technology (IT) and Information and Communications Technology (ICT). These data is gathered from first year students between for the entire academic year 2015-2016. The data being used for this research is based from previous study by Cantara, et.al <sup>[11]</sup>. Figure 1 shows the framework on the process of the approach.

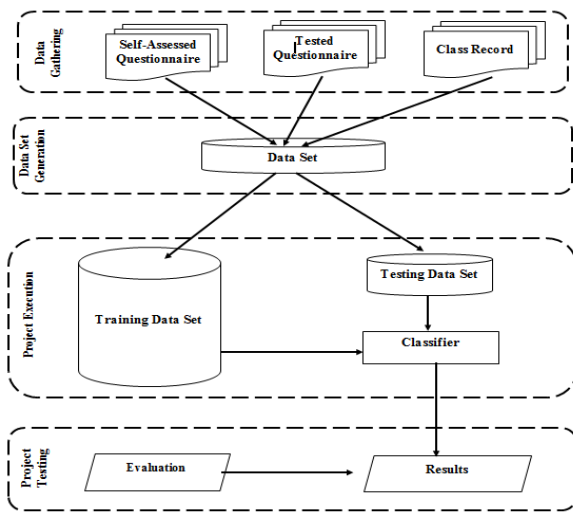


Figure 1. Framework of Process flow

### 3.1. Gathering of Data

The dependent variable considered is the final grades of the respondents in their programming subject (CS1101, IT1101 and ICT110). The final grade constitutes the academic performance of the students. The final grade ranges from 1.0 – 3.0 with remarks “Passed” and 5.0 with remarks “Fail”. With this seemingly numeric data, it is then divided into quartiles where the 4<sup>th</sup> quartile is the class for respondents with final grade of 5.0. The range 1.0 – 3.0, is then divided to the three quartiles: Q1: 1.0 – 1.6; Q2: 1.7 – 2.3; Q3: 2.4 – 3.0. Having all the final grades classified into different quartiles makes the academic performance a discrete variable i.e.  $C_1, C_2, C_3$  and  $C_4$ , rather than continuous. Although IQ has a numerical value, there is already an existing categorization for the range of IQ most commonly used by psychologists like Wechsler Scale and Stanford-Binet Scale [12][13][14]. Study habits, intrapersonal and interpersonal scores are taken from self-assessed surveys or questionnaires. A study habit is classified depending on the decile rank of their score. The survey items for interpersonal were validated with 0.863 alpha. The intrapersonal and interpersonal motivation factors are categorized based on their overall average from their responses.

There are 189 respondents will all the features available. This is divided into 70:30 where 70% of the data goes to the training set, while 30% of the data goes to the testing set. The training set size is 132 and the testing set is 57. Table 1 lists the elements of each set of variables.

Table 1. Variable Sets

Set	Elements
F (Feature set)	$\{f_w, f_x, f_y, f_z\}$
C (Classification Set)	$\{C_1, C_2, C_3, C_4\}$
W (IQ Classification set)	$\{w_1, w_2, w_3, w_4, w_5\}$
X (Study Habits Classification set)	$\{x_1, x_2, x_3, x_4, x_5\}$
Y (Intrapersonal Classification set)	Average $\{y_1, y_2, y_3, y_4, y_5\}$
Z (Interpersonal Classification set)	Average $\{z_1, z_2, z_3, z_4, z_5, z_6\}$

The feature set  $F$  is the independent variable which is used to predict classification set  $C$ . The feature set  $F$  has 4 subclass  $f_w$  for IQ,  $f_x$  for study habits,  $f_y$  for intrapersonal motivation, and  $f_z$  interpersonal motivation.

Table 2 is the overall frequency table for each  $F$  and  $C$  sets for the training set.

Table 2. Frequency table for the Training Set

$F \backslash C$	$C_1$	$C_2$	$C_3$	$C_4$
$z_1$	0	0	0	0
$z_2$	0	1	0	0
$z_3$	5	13	22	25
$z_4$	13	15	19	17
$z_5$	0	1	1	0
$z_6$	0	0	0	0
$y_1$	0	3	2	1
$y_2$	13	14	19	29
$y_3$	4	13	20	12
$y_4$	1	0	1	0
$w_1$	3	2	1	1
$w_2$	6	9	8	2
$w_3$	9	16	28	37
$w_4$	0	3	5	2
$x_5$	7	13	16	15
$x_1$	1	4	7	3
$x_2$	1	2	3	8
$x_3$	3	6	8	6
$x_4$	6	5	8	10

### 3.2 Treatment of Data

Naïve Bayes algorithm is being used to obtain the probabilities of the classification. Naïve Bayes uses the formula:

$$P(C_j | f_1, f_2, \dots, f_n) = P(C_j) \prod_{i=1}^n P(f_i | C_j)$$

where  $C_j \in C \forall j$  and  $f_n \in F$

(1)

Naïve Bayes algorithm has been used for many classification and clustering challenges. Naïve Bayes algorithm has been used in text classification, network traffic classification and even recommendation prediction. Although usually paired with data mining or educational data mining, features are just mined from the education database of the school like presented by Ahmand, et. al [8], Affendey et. al. [9], Sundar [10] and Durairaj and Vijitha [15]. For each  $C_j$  in  $C$ , 5 probabilities are needed are will be taken from training set table of frequencies in Table 2.

## 4. RESULTS

Measurement of the performance of the algorithm is the following:

$$Accuracy = \frac{\text{Number of Correctly Classified Entries}}{\text{Total number of Entries}}$$

(2)

The exact match accuracy of the algorithm is 45.61%. This is based on the Table 1

Table 1. Exact match accuracy table.

	Number of Entries	Percentage
Correctly Classified	26	45.61%.
Incorrectly Classified	31	54.39%

Below is the breakdown of the exact match accuracy per classification.

Table 2. Breakdown of Exact match accuracy

Class	Number of Correctly Classified Entries	Percentage
C <sub>1</sub>	2/7	28.57%
C <sub>2</sub>	3/13	23.07%
C <sub>3</sub>	13/22	59.09%
C <sub>4</sub>	8/15	53.33%

However, it must be noted that the classification is a range. For example, the range of C<sub>1</sub> is from 1.0 – 1.6, where 1.6 and 1.7 is very near each other and 1.7 belongs to C<sub>2</sub>. This is the accuracy of the algorithm wherein it predicted classification is a neighbor of the actual classification. Each classification is paired according to its nearest neighbor classification. C<sub>1</sub> is paired with C<sub>2</sub>, C<sub>3</sub> is paired with C<sub>4</sub>. C<sub>2</sub> is also paired with C<sub>3</sub>.

Table 3. Class-Pair accuracy

Class Pair	Number of Correctly Classified Entries	Percentage
C <sub>1</sub> , C <sub>2</sub>	6/20	30.00%
C <sub>2</sub> , C <sub>3</sub>	25/35	71.34%
C <sub>3</sub> , C <sub>4</sub>	28/37	75.68%

Lastly, in the objective of identifying that the student possess a higher possibility to pass, Table 4 shows the accuracy narrowing the classification into “Pass” or “Fail” where C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> are considered “Pass” and C<sub>4</sub> is considered “Fail”.

Table 4. Sensitivity and Specificity table

Classification	Number of Correctly Classified Entries	Percentage
Pass	32/42	76.18%
Fail	8/15	53.33%

### 5. CONCLUSION

In this paper, Naïve Bayes algorithm is used to predict the classification of the student’s academic performance. It is important to note that once the factors are available either during admission or within the first months of entry, the administrators of the institution can set expectation to the students. This allows the student to seek intervention to improve positive academic perform or avoid the negative academic performance. This will also set the expectation of the parents as it will allow support to the students and affect their perception about their academic performance [18][19].

C<sub>3</sub> can be considered as a borderline class between “Pass” and “Fail” since its range is between 2.3 – 3.0. A great effort can be done to intervene with the students classified under C<sub>3</sub>. Needless to say, greater effort is given to those classified as C<sub>4</sub> because this is the class for “Fail”.

Future work may include improving the algorithm to obtain higher accuracy ratings. Another recommendation may include creating an application which would automate the recommendation and decision for a student’s admission and possible intervention.

### 6. REFERENCES

- [1] Ping, C.-M. (2005). Prediction of Student Academic Performance with Psychological Constructs beyond Academic Skills. *Journal of Humanities and Social Sciences*, 1(1), 83-89.
- [2] Garton, B. L., Dyer, J. E., & King, B. O. (2000). The Use of Learning Styles and Admission Criteria in Predicting Academic Performance and Retention of College Freshmen. *Journal of Agricultural Education*, 41(2), 46-53.
- [3] Kruck, S. E., & Lending, D. (2003). Predicting Academic Performance in an Introductory College-Level IS Course. *Information Technology, Learning, and Performance Journal*, 21(2), 9-15.
- [4] Brazdău, O., & Mihai, C. (2011). The Consciousness Quotient: A New Predictor of the Students’ Academic Performance. *Elsevier Procedia Social and Behavioral Sciences*, 11, 245–250.
- [5] Verkuyten, M., Thijs, J., & Canatan, K. (2001). Achievement Motivation and Academic Performance Among Turkish Early and Young Adolescents in the Netherlands. *Genetic, Social, and General Psychology Monographs*, 127(4), 378–408.
- [6] Ting, S. R. (2001). Predicting Academic Success of First-Year Engineering Students from Standardized Test Scores and Psychosocial Variables. *International Journal of Engineering Education*, 17(1), 78-80.
- [7] Wang, R., Harari, G., Hao, P., Zhou, X., & Campbell, A. T. (2015). Smar tGPA: How Smar tphones Can Assess and Predict Academic Performance of Colleg e Students. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 295-306). Osaka, Japan: UbiComp ’15, ACM.
- [8] Ahmad, F., Ismail, N., & Aziz, A. (2015). The Prediction of Students’ Academic Performance Using Classification Data Mining Techniques. *Applied Mathematical Sciences*, 9(129), 6415 - 6426.
- [9] Affendey, L., Paris, I., Mustapha, N., & Muda, Z. (2010). Ranking of Influencing Factors in Predicting Students Academic Performance. *Information Technology Journal*, 9(4), 832-837.
- [10] Sundar, P. (2013). A Comparative Study for Predicting Students Academic Performance Using Bayesian Network Classifiers. *IOSR Journal of Engineering*, 3(2), 37-42.
- [11] Cantara, A. D., Capao, K. A., Ceniza, A. M., Mendoza , E. P., Polinar, S. B., Tero, J. M., & Sabellano, M. G. (2016). The Study on the Correlation amongst the Students’ Final Grade, Study Habits, Behavioral, Intellectual, and Environmental Factors Using Linear Regression. *University of San Carlos*.
- [12] Scheiman, M., & Rouse, M. W. (2016). *Optometric Management of Learning-related Vision Problems*. St. Louis, MO: Elsevier.
- [13] Yudofsky, S. C., & Hales, R. E. (2007). *American Psychiatric Press Textbook of Neuropsychiatry* (5th ed.). Washington, DC: American Psychiatric Press.
- [14] Skuy, M., Taylor, M., O’Carroll, S., Fridjhon, P., & Rosenthal, L. (2000). Performance of black and white South African children on the Wechsler Intelligence Scale for children-revised and the Kaufman assessment battery. *Psychological Reports*, 86(3), 727-737.
- [15] Durairaj, M., & Vijitha, C. (2014). Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), 5987-5991.

- [16] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110.
- [17] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.
- [18] Grossman, J. A., Kuhn-McKearin, M., & Strein, W. (2011). Grossman, J. A., Kuhn-McKearin, M., & Strein, W. (2011). Parental Expectations and Academic Achievement: Mediators and School Effects. *Annual Convention of the American Psychological Association*, 4. Washington, DC.
- [19] Topor, D. R., Keane, S. P., Shelton, T. L., & Calkins, S. D. (2010). Parent involvement and student academic performance: A multiple mediational analysis. *Journal of prevention & intervention in the community*, 38(3), 183-197.