Artificial System for Prediction of Student’s Academic Success from Tertiary Level in Bangladesh

By Linkon Chowdhury & Shahana Yeasmin

BGC Trust University Bangladesh

Abstract - Every year a large scale of students in Bangladesh enrol in different Universities in order to pursue higher studies. With the aim to build up a prosperous career these students begin their academic phase at the University with great expectation and enthusiasm. However among all these enthusiastic and hopeful bright students many seem to become successful in their academic career and found to pursue the higher education beyond the undergraduate level. The main purpose of this research is to develop a dynamic academic success prediction model for universities, institutes and colleges. In this work, we first apply chi square test to separate factors such as gender, financial condition and dropping year to classify the successful from unsuccessful students. The main purpose of applying it is feature selection to data. Degree of freedom is used to P-value (Probability value) for best predictors of dependent variable. Then we have classify the data using the latest data mining technique Support Vector Machines(SVM). SVM helped the data set to be properly design and manipulated. After being processed data, we used the MATH LAB for depiction of resultant data into figure. After being separation of factors we have had examined by using data mining techniques Classification and Regression Tree (CART) and Bayes theorem using knowledge base. Proposition logic is used for designing knowledge base. Bayes theorem will perform the prediction by collecting the information from knowledge Base. Here we have considered most important factors to classify the successful students over unsuccessful students are gender, financial condition and dropping year. We also consider the sociodemographic variables such as age, gender, ethnicity, education, work status, and disability and study environment that may influence persistence or academic success of students at university level. We have collected real data from Chittagong University Bangladesh from numerous students. Finally, by mining the data, the most important factors for student success and a profile of the typical successful and unsuccessful students are identified.

Keywords : Chi Value, P-value, CART, SVM, Bayes theorem, Degree of freedom.

GJCST-C Classification : I.2.0

Strictly as per the compliance and regulations of:

© 2012. Linkon Chowdhury & Shahana Yeasmin. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Artificial System for Prediction of Student’s Academic Success from Tertiary Level in Bangladesh

Linkon Chowdhury & Shahana Yeasmin

Abstract - Every year a large scale of students in Bangladesh enroll in different Universities in order to pursue higher studies. With the aim to build up a prosperous career these students begin their academic phase at the University with great expectation and enthusiasm. However among all these enthusiastic and hopeful bright students many seem to become successful in their academic career and found to pursue the higher education beyond the undergraduate level. The main purpose of this research is to develop a dynamic academic success prediction model for universities, institutes and colleges. In this work, we first apply chi square test to separate factors such as gender, financial condition and dropping year to classify the successful from unsuccessful students. The main purpose of applying it is feature selection to data. Degree of freedom is used to P-value (Probability value) for best predictors of dependent variable. Then we have classify the data using the latest data mining technique Support Vector Machines (SVM). SVM helped the data set to be properly design and manipulated. After being processed data, we used the MATH LAB for depiction of resultant data into figure. After being separation of factors we have had examined by using data mining techniques Classification and Regression Tree (CART) and Bayes theorem using knowledge base. Proposition logic is used for designing knowledge base. Bayes theorem will perform the prediction by collecting the information from knowledge Base. Here we have considered most important factors to classify the successful students over unsuccessful students are gender, financial condition and dropping year. We also consider the sociodemographic variables such as age, gender, ethnicity, education, work status, and disability and study environment that may influence persistence or academic success of students at university level. We have collected real data from Chittagong University Bangladesh from numerous students. Finally, by mining the data, the most important factors for student success and a profile of the typical successful and unsuccessful students are identified.

Keywords : Chi Value, P-value, CART, SVM, Bayes theorem, Degree of freedom.

1. INTRODUCTION

Student retention is an indicator of academic performance and enrolment management of the university. Increasing student retention or persistence is a long term goal in all academic institution. High rates of student attrition have been a concern for the past several years at the Chittagong University Institute of Computer Science and Engineering. Level of retention rate goes higher by various reasons and one the most important reasons are government funding, scholarship in the tertiary education environment. There are both academic and administrative staff are under pressure to come up with strategies that could increase retention rates on their courses and programs. The lowest student retention rates at all institutions of higher education are first-year students, who are at greatest risk of dropping out in the first term or semester of study or not completing their programs or degree. Therefore most retention studies address the retention of first-year students. Consequently, the early identification of vulnerable students who are prone to drop their courses is crucial for the success of any retention strategy. This would allow educational institutions to undertake timely and pro-active measures. Once identified, these at risk students can be then targeted with academic and administrative support to increase their chance of staying on the course.

Data from 2004 to 2010, covering over 200 enrolled students stored in the Computer Science and Engineering was used to perform a quantitative analysis of study outcome. There were three main types of objectives conducted in this study. The first approach is descriptive which is concerned with the nature of the dataset such as the frequency table and the relationship between the attributes obtained using cross tabulation analysis (contingency tables). In addition, feature selection is conducted to determine the importance of the prediction variables for modelling study outcome. The third type of data mining approach, i.e. predictive data mining is conducted by using two different types of classification trees. The classification tree models have some advantages. Secondly, the classification tree models are non-parametric and can capture nonlinear relationships and complex interactions between predictors and dependent variable. We decided not to use other data mining techniques such as neural networks and support vector machines even though in some cases they could achieve higher accuracy, because their structure is not transparent and usually described as a black box. It is also difficult to explain their results and how they work to a user who would like
to apply them to a new set of data. Finally, a comparison between these models was conducted to determine the best model for the dataset. The population of this study was entering CSE in the first -year from 2004 through 2009. These were full-time degree-seeking students. Entering in the first-year session were excluded from the study for a few reasons. The independent or predictor variables fell into three main categories. These were the students' personal information, high school background. The personal information recorded for each student was:

- Ethnicity (Bengali, Marma, Chakma and Others)
- Gender
- Age
- Financial Status
- Work Status
- Disability

II. Collected Statistical Data

As part of the data-understanding phase we carried out the data on the table 1 and table 2. The Table 2 reports the results. Based on the results shown majority of Information Systems students are female (over 38%). However, percentage of female students who successfully complete the course are higher (41%) which suggests that female students are more likely to[3] pass the course than their male counterpart. When it comes to age over 26% of students are above 24. This age group is also more likely to fail the course because their percentage of students who failed the course in this age group (11.9%) is higher than their overall participation in the student population (26.2%). Statistical data on 42 students:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Domain Name</th>
<th>Count</th>
<th>Total</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>26</td>
<td>61.9</td>
<td>58.6</td>
<td>69.2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>16</td>
<td>38.1</td>
<td>41.4</td>
<td>30.8</td>
</tr>
<tr>
<td>Age Group</td>
<td>&gt;24</td>
<td>11</td>
<td>26.2</td>
<td>20.7</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>&lt;=24</td>
<td>31</td>
<td>73.8</td>
<td>79.3</td>
<td>61.5</td>
</tr>
<tr>
<td>Disabilities</td>
<td>Yes</td>
<td>1</td>
<td>2.4</td>
<td>0.0</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>41</td>
<td>97.6</td>
<td>100.0</td>
<td>92.3</td>
</tr>
<tr>
<td>Financial Support</td>
<td>Yes</td>
<td>23</td>
<td>54.8</td>
<td>62.1</td>
<td>38.5</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>19</td>
<td>45.2</td>
<td>37.9</td>
<td>61.5</td>
</tr>
</tbody>
</table>

Students with it are more likely to fail than those without it. There are huge differences in percentage of students who successfully completed [4] the course depending on their ethnic origin. A substantial number of students (over 55%) have financial support more vulnerable than the other two categories in this variable.

III. Support Vector Machines

Support Vector Machine (SVM) is one of the latest clustering techniques which enables machine learning concepts to amplify predictive accuracy in the case of axiomatically diverting data those are not fit properly. It uses inference space of linear functions in a high amplitude feature space, trained with a learning algorithm. It works by finding a hyper plane that linearly separates the training points, in a way such that each resulting subspace contains only points which are very similar. First and foremost idea behind Support Vector Machines (SVMs) is that it constituted by set of similar supervised learning. An unknown tuple is labeled with the group of the points that fall in the same subspace as the tuple. Earlier SVM was used for Natural Image processing System (NIPS) but now it becomes very popular is an active part of the machine learning research around the world. It is also being used for pattern classification and regression based applications. The foundations of Support Vector Machines (SVM) have been developed by V. Vapnik.

SVM is very effective in various data and information classification process. An expert should bear in mind two important factors for implementing SVM, these two factors or techniques are mathematical programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured.
We can calculate the weight boundary maximum margin by using the following equation:

\[
\text{margin} = \arg \min_{x \in D} d(x) = \arg \min_{x \in D} \frac{|x \cdot w + b|}{\sum_{i=1}^{d} w_i^2}
\]

Another interesting question is why maximum margin? There are some good explanations which include better empirical performance. Another reason is that even if we’ve made a small error in the location of the boundary this gives us least chance of causing a misclassification. The other advantage would be avoiding local minima and better classification. The goals of SVM are separating the data with hyper plane and extend this to non-linear boundaries using kernel trick [8] [11]. For calculating the SVM we see that the goal is to correctly classify all the data. For mathematical calculations we have,

[a] If \( Y_i = +1 \); \( wx_i + b \geq 1 \)

[b] If \( Y_i = -1 \); \( wx_i + b \leq 1 \)

[c] For all \( i \); \( y_i (wx_i + b) \geq 1 \)

In this equation \( x \) is a vector point and \( w \) is weight and is also a vector. So to separate the data [a] should always be greater than zero. Among all possible hyper planes, SVM selects the one where the distance of hyper plane is as large as possible. If the training data is good and every test vector is located in radius \( r \) from training vector. Now if the chosen hyper plane is located at the farthest possible from the data [12]. This desired hyper plane which maximizes the margin also bisects the lines between closest points on convex hull of the two datasets. Thus we have [a], [b] & [c].

**IV. Classification and Regression Tree (CART)**

Classification and Regression Tree (CART) has many advantages over classification methods. It is naturally non-parametric. CART can handle [5][6] numerical data that are highly skewed. It eliminates analyst time, which would otherwise be spent determining whether variables are normally distributed. CART identifies “splitting” variables based on a complete search of all possibilities. CART is able to search all possible variable splitters, even in problems with many hundreds of possible predictors as well as dealing missing variable. Finally, CART trees are moderately simple and non-statisticians.

The purpose of an analysis based on classification tree is to find out the factors that contribute the separation of successful and unsuccessful students among the all recorded data. When the classification tree is generated it is possible to calculate the probability of each student’s outcome. If once the classification tree is formed, it is possible to predict the study outcome for newly enrolled student. In each tree node the percentages for successful and unsuccessful students is given and also absolute size of the node. The variable names above the node are the predictor’s criteria that make the split for the node according to classification and regression tree. Each node is split according to predictor criteria. The searching is stops when the split with the largest improvement in goodness of fit. Possible predictor variables in the dataset were included in the classification tree in splits process, which is detected in feature selection criteria. We used two stopping criteria in the training process:
1. A maximum tree depth has been proceeding into 3 levels for CHAID tree.
2. A maximum tree depth has been proceeding into 4 levels for CART tree.

Lastly for each classification tree we have assigned different costs to the classification outcomes i.e.: classification matrix. It is one of the processes of increasing the correctly classified student’s outcomes. We categories the student’s outcome according to Pass and Fail in the academic courses.

V. Our Contribution

In this research we explore the concepts and technique of SVM to classify the data collected in our experiments. We have approximately collected twelve hundreds (1200) data from University of Chittagong where about thirty (30000) thousands students are studying. To assess these large amounts of data we have found that SVM is very efficient and exact technique in our proceedings. By imposing the SVM, we have mapped the data to meaning full forty two (42) data which are shown in the figure 3 and 4. In figure 3 we have depicts that the reasons for the age where mainly focused on the pivotal age of greater or less than twenty four.

2. Calculating chi square value for each cell using the following formula

\[ \chi^2 = \sum_{i=1}^{n} \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}} \]

Step 1: First insert the observed value in each cell of observable table. Inserted value collected from record.

<table>
<thead>
<tr>
<th>Domain category</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>Category 2</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td>a + b</td>
<td>b</td>
<td>a + b + c</td>
</tr>
</tbody>
</table>

Step 2: Calculate expected value for every cell of the describing table.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>a1 = (a+b)/(a+b+c)</td>
<td>b1 = (a+b)/(a+b+c)</td>
<td>a1 + b1</td>
</tr>
<tr>
<td>Category 2</td>
<td>c1 = (b+d)/(a+b+c)</td>
<td>d1 = (b+d)/(a+b+c)</td>
<td>c1 + d1</td>
</tr>
<tr>
<td>Total</td>
<td>a1 + c1</td>
<td>b1 + d1</td>
<td>a1 + b1</td>
</tr>
</tbody>
</table>

Step 3: Calculating degree of freedom using rule

\[ \text{df} = (\text{No.of.rows}-1) \times (\text{No.of.columns}-1) \]

Step 4: Calculate p-value (probability value) using following method in Ms Excel

\[ \text{P-value} = \text{CHIDIS} (\text{Chi value, df}) \]
VI. EXPLANATION OF CHI-SQUARE (χ²) AND P-VALUE

Step 1: consider the domain is financial support in which category1=Yes, category2=No, Option1=Pass and Option2=Fail. The value of every cell collects from database.

<table>
<thead>
<tr>
<th>Financial Support</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>18</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>No</td>
<td>11</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>13</td>
<td>42</td>
</tr>
</tbody>
</table>

Step 2: calculating expected value for each cell using describing formula

<table>
<thead>
<tr>
<th>Financial Support</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>23*29/42=15.88</td>
<td>23*13/42=7.12</td>
<td>23</td>
</tr>
<tr>
<td>No</td>
<td>19*29/42=13.12</td>
<td>19*13/42=5.88</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>13</td>
<td>42</td>
</tr>
</tbody>
</table>

Step 3: calculating chi value for every cell using the describing formula:

<table>
<thead>
<tr>
<th>Financial Support</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>χ²=(18-15.88)²/15.88 = 0.283</td>
<td>χ²=(5-7.12)²/7.12 = 0.631</td>
</tr>
<tr>
<td>No</td>
<td>χ²=(11-13.12)²/13.12 = 0.343</td>
<td>χ²=(8-5.88)²/5.88 = 0.764</td>
</tr>
</tbody>
</table>

Step 4: calculate total chi -value for domain

χ² = 0.283 + 0.631 + 0.343 + 0.764 = 2.021

Overall chi-value for every domain in Pass and Fail Category:

Table 3: Individual Chi-value for each category

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Possible category</th>
<th>Pass</th>
<th>Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Support</td>
<td>Yes</td>
<td>0.283</td>
<td>0.631</td>
</tr>
<tr>
<td>Age group</td>
<td>Age&gt;24</td>
<td>0.337</td>
<td>0.753</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>0.050</td>
<td>0.112</td>
</tr>
<tr>
<td>Disabilities</td>
<td>Yes</td>
<td>0</td>
<td>1.54</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>0.082</td>
<td>0.182</td>
</tr>
</tbody>
</table>

Step 5: calculating degree of freedom using following the rule

Degree of freedom df = (2-1)*(2-1) = 1

Step 6: calculate p-value (probability value) using in Ms Excels

P-value = CHIDIS (2.021, 1) = 0.155

VII. FACTORS SELECTION

Factors selection is an important process to assess the prediction of dropout of the students. The prediction has relate the variable that determines the rate of success. The number of predictor variables is not so large and we don’t have to select the subset of variables for further analysis which is the main purpose of applying feature selection to data. However, feature selection could be also used as a pre-processor for predictive data mining to rank predictors according to the strength of their relationship with dependent or outcome variable. During the factors selection process no specific form of relationship, neither linear nor nonlinear, is assumed. The outcome of the factors selection would be a rank list of predictors according to their importance for further analysis of the dependent variable with the other methods for regression and classification. Here the figure below shows the relative outcome of the predictor’s value.

Figure 5: Importance plot for predictors

Results of factors selection has presented in Figure 3 on the importance plot and in Table 1. The top three predictors for the study outcome are financial support of students, disabilities, age group and gender in which they are study.

Table 4: Best predictors for dependent variable

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Chi-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Support</td>
<td>2.021</td>
<td>.155</td>
</tr>
<tr>
<td>Disabilities</td>
<td>1.6</td>
<td>.206</td>
</tr>
<tr>
<td>Age Group</td>
<td>1.477</td>
<td>.224</td>
</tr>
<tr>
<td>Gender</td>
<td>.426</td>
<td>.514</td>
</tr>
</tbody>
</table>
In all three cases, i.e. for all three definitions of the dependent variable, if the top 4 variables are selected, we get the same list of predictors. Therefore we can conclude that the list of important predictors is quite robust to changes in the study outcome definition. We may proceed into the next step using the top 4 variables:

1. Financial Support
2. Age Group
3. Gender
4. Disabilities

From Table 4, P-values from the last column only the first three chi-square values are significant at 10% level. Though the results of the feature selection suggested continuing analysis with only the subset of predictors, which includes Financial Support, Age group and Gender, we have included all available predictors in our classification tree analysis. We follow an advice given in Luan & Zhao (2006) who suggested that even though some variables may have little significance to the overall prediction outcome, they can be essential to a specific record.

**a) Contribution of Cart**

Figure 6 shows the CART classification tree for study outcome. It shows that only three variables were used to construct the tree: (1) financial support (2) age group and (3) gender.

The largest successful group (i.e. students who successfully completed the course) consists of 18 (78.76% of all participants) students (Node 1). The financial support of students in this group is yes. Students in this group enrolled on the age group in either age>24 or age<=24. The largest successful group (i.e. age group) contains 9 (60% of all participants) students (Node 4).

**VIII. Knowledge Base for Collected Data**

A knowledge base in artificial intelligence is a place where information are stored or designed for machine or device by which it will work. In general, a knowledge base is a consolidate stock for information: a library, a database of related information about a particular subject could all be considered to be examples of knowledge bases. The process of building knowledge base is called knowledge engineering. A knowledge base is integrated collection of choosing logic, building a knowledge base, implementing [31] the proof theory, inferring new facts. The main advantage of engineering is that it requires less commitment and thus less work. To help the focus the development of knowledge base and to integrate the designer’s thinking the following five step methodology can be used:

1. Decide what to talk about
2. Decide on a vocabulary of predicates, function, and constant.
3. Encode general knowledge about the domain.
4. Encode a description of the specific problem instance.
5. Pose queries to the inference procedure and answers.

In our work we have described a simple method of probabilistic inference that is, the computation from observed evidence of posterior probabilities for query propositions. We have used the joint probability as the knowledge base from which answer to all question may be derived. We have had built the knowledge base by considering two Boolean variables. The table 7 is an example of two valued propositional logic which is the bases of knowledge base representation:

**Table 7: Concepts of propositional logic to design a Knowledge Base using the proposition of Boolean events A, B and C.**

<table>
<thead>
<tr>
<th>B</th>
<th>¬B</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>¬C</td>
</tr>
<tr>
<td>¬A</td>
<td>001</td>
</tr>
</tbody>
</table>

Based on table 7, we have designed the knowledge base (Joint probability distribution) for our research activity. Here we have considered those events which have true (one or 1) Boolean values. Table 8 is an example of knowledge base for events A, B and C:
By keeping the similarities s with the table 5, we compared our factors as financially good and financially not good, fail and not fail and so on. The designing of knowledge base for the factors which we are considered are given in table 6:

Table 6: Fully join probability distribution (Knowledge base)

<table>
<thead>
<tr>
<th>Financial</th>
<th>Age&lt;=$24</th>
<th>Age&gt;$24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>0.125</td>
<td>0.044</td>
</tr>
<tr>
<td>Pass</td>
<td>0.279</td>
<td>0.099</td>
</tr>
</tbody>
</table>

Where
0.125+0.044+0.103+0.037+0.279+0.099+0.231+0.082=1

IX. BAYES THEOREM AND CONDITIONAL PROBABILITY

Bayes’ theorem and conditional probability are opposite to each other. Given two dependent events A and B. The conditional probability of P (A and B) or P (B/A) will be P (A and B)/P (A). Related to this formula a rule is developed by the English Presbyterian minister Thomas Bayes (1702-61). According to the Bayes rule it is possible to determine the various probabilities of the first event given the outcome of the second event in a sequence of two events.

The conditional probability:

$$P(B/A) = \frac{P(A \land B)}{P(A)}$$  \hspace{1cm} (1)

The equation (1) will help to find out the probabilities of B after being occurrences of the A, we get the Bayes’ theorem for these two events as follows:

$$P(A/B) = \frac{P(A \land B)}{P(B)}$$  \hspace{1cm} (2)

If there are more events like A1, A2, and B1, B2. In this case the Bayes theorem to determine the probability of A1 based on B1 will be as follows:

$$P(A1/B1) = \frac{P(A1 \land B1)}{P(B1)} = \frac{P(A1) \cdot P(B1/A1)}{P(A1) \cdot P(B1/A1) + P(A2) \cdot P(B2/A2)}$$

X. EXPERIMENTAL RESULT

Consideration Classification and Regression Tree (CART):

For Node 4:
IF Financial Support = "Yes" AND Age<=$24
 THEN "PASS" = (18/23)*9/15 = 0.49

Now applying the Bayes theorem on table 5 we have got the following outcomes:

IF one student pass based on his financial condition is "Yes" and age<=$24 then
P (Pass | Financial condition = "Yes" \^ age<=$24) =
P (Pass \land Financial condition = "Yes" \^ age<=$24) / P (Financial condition = "Yes" \^ age<=$24)
=0.279
P (Financial condition = "Yes" \^ age<=$24) =0.125+0.044 =0.404

P (Pass | Financial condition = "Yes" \^ age<=$24) =0.125+0.044 =0.404

The total resultant of Bayes Theorem and CART of all data considering financial condition and age group we have got the following table 7:

Table 7: Bayes Rules to predict the academic success

<table>
<thead>
<tr>
<th>Rule</th>
<th>Study Outcome</th>
<th>Probability CART</th>
<th>Probability Bayes Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF Financial Support = &quot;Yes&quot; AND Age&lt;=$24 THEN</td>
<td>PASS</td>
<td>0.47</td>
<td>0.69</td>
</tr>
<tr>
<td>IF Financial Support = &quot;Yes&quot; AND Age&gt;$24</td>
<td>PASS</td>
<td>0.39</td>
<td>0.69</td>
</tr>
<tr>
<td>IF Financial Support = &quot;NO&quot; AND Age&lt;=$24</td>
<td>PASS</td>
<td>0.35</td>
<td>0.68</td>
</tr>
<tr>
<td>IF Financial Support = &quot;NO&quot; AND Age&gt;$24</td>
<td>PASS</td>
<td>0.29</td>
<td>0.68</td>
</tr>
</tbody>
</table>
XI. Conclusion

This study examines the background information from enrolment data that impacts upon the study outcome of Information Systems students at the Department of Computer Science & Engineering. Based on results from feature selection (Figure 5 and Table 4), the CART trees presentation it was found that the most important factors that help separate successful from unsuccessful students are financial support, age group and gender. Demographic data such as gender and age though significantly related to the study outcome, according to the feature selection result.

Based on results from table 5 and 6 by implementing the knowledge of propositional knowledge base and Bayes theorem based on knowledge base to predict the academic success. Better result is found from using Bayes Network.

This study is limited in three main ways that future research can perhaps address. Firstly, this research is based on background information only. Secondly, we used a dichotomous variable for the study outcome with only two categories: pass and fail. Thirdly, this study is limited in three main ways that future research can perhaps address. Firstly, this research is based on background information only.

References Références Referencias


than their peers? *Research in Rural Education, 6*(1), 31-34.


