

# Classifying Student's Engagement in Computer Games using Linear Discriminant Analysis

Kennet G. Cuarteros<sup>1</sup>, Rose May L. Puerte<sup>2</sup>

<sup>1</sup>University of Science and Technology of Southern Philippines, Lapasan Highway, Cagayan de Oro City, Philippines

<sup>2</sup>University of Science and Technology of Southern Philippines, Lapasan Highway, Cagayan de Oro City, Philippines

**Abstract:** *Linear Discriminant Analysis can be used to determine which variable discriminates between two or more classes, and to derive a classification model for predicting the group membership of new observations. For each of the groups, LDA assumes the explanatory variables to be normally distributed with equal covariance matrices. The simplest LDA has two groups. To discriminate between them, a linear discriminant function that passes through the centroids of the two groups can be used. The study used Linear Discriminant Analysis in classifying student as addicted or non-addicted in computer games. The study conducted a survey in the form of questionnaire to the students who are playing computer games and a student in Mindanao University of Science and Technology (MUST). Young Diagnostic Test (eight-item questionnaire) was adopted and used Likert Scale to answer the survey questionnaire. The researcher was able to classify 100 students by using Linear Discriminant Analysis. It was found out that 61 out of 63 or 96.83% is correctly classified as non-addicted and 35 out of 37 or 94.59% is correctly classified as addicted to computer games. Moreover, the study has 4.29% of average misclassification probability which implies that the Linear Discriminant Analysis performs better in classifying behavioral addiction. The study further showed that the students in MUST can manage their time properly as to when to study and when to play computer games as part of their recreational past time.*

**Keywords:** linear discriminant analysis, students' engagement, computer games

## 1. Introduction

Addiction (termed substance dependence by the American Psychiatric Association) is defined as a maladaptive pattern of substance use leading to clinically significant impairment or distress, as manifested by three (or more) of the following: tolerance, withdrawal, the substance is often taken in larger amounts or over a longer period than intended, there is a persistent desire or unsuccessful efforts to cut down or control substance use and the like.

A clinical psychologist for TechAddiction Dr. Brent Conrad (2013) says that computer game addiction generally refers to an excessive, unhealthy amount of playing computer games. Rather than engaging in the real world, an addicted user devotes the majority of his or her time to gaming. The addicted gamer often isolates him/herself from others, ignores more important responsibilities, and is often obsessed with obtaining higher status / rankings / achievements in his/her favorite computer game.

College is one of the most crucial stages of education because during this stage students are taught to be independent in life. This is also the time that the children are forcedly moved to urban area in order to take college whereas they can have a new freedom from parental controls. Because of the fast advancement of technology nowadays, students were getting involved with the use of computer. There are a lot of computer shops have been established in the city and it is widely used for different purposes. Some they use this for education purpose and some for pleasure. The critical problem is that the students use the computer for wrong purposes; they spent so much time playing computer games that they ignore their study or other important activities. Students tend to play games first, browse websites

and make new friends with the use of the internet rather than studying. Psychological and Emotional, Health, Relationship with the family, Academic and Social are some of the main areas that are often affected by computer game addiction.

Technological addictions are operationally defined as non-chemical (behavioral) addictions that involve human-machine interaction. They can either be passive (e.g., television) or active (e.g., computer games), and usually contain inducing and reinforcing features which may contribute to the promotion of addictive tendencies. Technological addictions can be viewed as a subset of behavioral addiction and feature core components of addiction, (i.e., salience, mood modification, tolerance, withdrawal, conflict and relapse; see Griffiths, 2006).

Despite the differences between seemingly disparate addictions, it's easy to see that there are common patterns to the addictions themselves, as well as common patterns in the addicts' backgrounds. (Langley, 2010). Computer Games Addiction was similar to other behavioral addiction like compulsive gambling or commonly known as gambling addiction, that is, the players felt satisfied when they won games. Assessments for problematic gambling are often used to tests for computer/internet addiction due to similarities between the two are both behavioral based addiction. (Brenner 1996; Young 1996a; Poon 2012).

A number of differing diagnostic criteria have been used in Internet addiction studies. One of the most commonly used criteria was the one used by Young (1996a) and subsequently by others. The diagnostic questionnaire consisted of eight items modified from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria for pathological gambling.

The present study intends to use Linear Discriminant Analysis to classify students as addicted and non-addicted to computer games. The study adopted the Young's Diagnostic Test for Internet Addiction that shows reliable and valid measure that has been used in many research about computer games addiction. The eight-item questionnaire measures the extent of client's involvement with the computer and classifies the addictive behavior. To assess the computer games addiction test, the survey questionnaire are answer based on the five Likert point, whereas, 1 indicates Strongly Disagree, 2 indicates Disagree, 3 indicates neither disagree nor agree, 4 indicates Agree and 5 indicates Strongly Agree.

## 2. Literature Review

### 2.1 Studies Related on Factors Affecting Computer Game Addiction

There is also a recent study conducted by Poon, A. (2012) entitled "Computer Game Addiction and Emotional Dependence". Computer games (games played on a computer, tablet, or any web-enabled device) have salient qualities, especially MMORPGs that cause addictive symptoms. The study investigated computer game addiction and usage in Trinity College students. A sample of 114 students ( $M = 20.4$  years of age, 61% female) was divided into a non-addicted, social player, and computer-addicted group based on Young's Diagnostic Questionnaire. The results showed that there is no significant correlation between day quality and computer game usage, and no significant correlation between emotional dependence and group. Not enough computer-game addicted subjects participated ( $n=4$ ), so it was unable to be concluded that computer game addicts have lower GPA, or prefer a certain type of game. However, the large number of social players ( $n=57$ ) shows that more research needs to be conducted to analyze the prevalence of addictive symptoms.

A study of Supaket, P. et al., (2008) entitled "Factors Affecting Computer Game Addiction and Mental Health of Male Adolescents in Mueang District, Si Sa Ket Province" This survey research aimed to study the factors affecting computer game addiction, mental health status, and the association of computer game addiction with the mental health of male adolescents in Mueang District, Si Sa Ket Province. The samples were 333 Mathayomsuksa One to Three students in Si Sa Ket municipal school selected by stratified random sampling to classify the schools and the grade. Data were collected by self-administered questionnaires and Thai General Health Questionnaire 28 (Thai GHQ 28) during August 28, 2007 to September 28, 2007 and analyzed by frequency, percentage, means, Chi-square test and Multiple logistic regression. The study found that factors with a statistically significant effect on the computer game addicted were the availability of Internet-connected PCs in house and the inclination of friends. Those adolescents with home internet had a higher chance of being addicted than those without home internet and those with higher friend inclination had a higher chance of being addicted than those with lower friend inclination. On the mental health status of the male adolescents, the study

revealed that adolescents who were addicted had poor mental health and computer game addiction was associated with weakened mental health status; a higher proportion of those who were addicted had poor mental health than those who were not addicted.

"An Analysis of Internet Addiction Levels of Individual According to Various Variables" by Sahin, C. (2011) states that the concept of internet addiction refers to the excessive use of internet which in turn causes various problems in individual, social and professional aspects. The aim of the study was to determine internet addiction levels of internet users from all age groups. The study used survey model. Study group of the study consisted of a total of 596 people from all age groups. "Personal Information Form" and "Internet Addiction Scale" were used for data collection. Arithmetic mean, standard deviation, independent sampling and t test, ANOVA and LSD tests were performed on collected data. The findings of the study revealed that the individuals had low levels of internet addiction both in sub-scales and in the general of the scale according to age groups. It was found that there was a significant difference between internet addiction scores of the individuals who belonged to the age group of 19 and below and 30 and below. There was a significant difference between the internet addiction scores of students and other professional groups. It was found that internet addiction levels of males were higher than those of females.

### 2.2 Studies Related to Classification of Computer Game Addiction

Tuisima et al. (2012) developed an agent program that collects the keyboard pressing and mouse clicking. Data was gathered in a log file stored in student's computer and set through the internet. Considering the several attributes used education level, gender, duration of playing game per day, duration of mouse clicking in game per day, duration of keyboard pressing in game per day, total number of mouse clicking and keyboard using in game playing per day, etc. Using Backpropagation Neural Networks they classified level of addiction based on their characteristic, into four categories: Long Term, Casual, Real Time and Turn-Based Games. Then compared the results to Decision Tree Learning and the accuracy of the obtained model was tested by the standard 10-fold Cross Validation approach.

### 2.3 Studies Related to Computer Addiction Problem

People with an addiction do not have control what they are doing. It may reach to a point at which they can be harmful to themselves and to others. Yeoh (2012) reported about a 23 year old guy who was found slumped over the keyboard in the front of his computer due to heart attack after apparently playing continuously for over 15 hours at a cybercafé. The guy has been at the cybercafé from 10pm on Sunday until 1pm the next day before returning home to his own computer. In other case, a three-month-old South Korean girl starved to death after her parents repeatedly left her alone for 12-hour stretches to rendezvous at a local Internet café. While online, the couple were raising a virtual daughter through a computer game known as "Prius Online."

seemingly as a form of escapism from their own lives.“ After returning in the café, the 41-year-old father and 25-year-old mother found their real child dead and it was revealed that the baby died because of prolonged malnutrition.

## 2.4 Studies Related on Classifying using Linear Discriminant Analysis

Discriminant Analysis is a technique that allows us to study differences between of cases and groups with respect to different predictor variables. One of these examples is the study of Casio (2012) entitled Development of Classification Models for Predicting Success of Electronics Engineering Graduates of Mindanao University of Science and Technology on licensure examination. The study classifies the success of the graduates by using Logistic Regression and Linear Discriminant Analysis. The study consists of twenty-six explanatory variables with two demographic factors, two pre-admission factors and twenty-two school-related factors. It was found out that the classification model produced by the Linear Discriminant Analysis is better in predicting success in the Electronics Engineering licensure examination than the classification model produced by the Logistic Regression. In other case, Cabangca (2011) used Linear Discriminant Analysis to classify as low or high risk borrower of one lending companies in the Philippines. The study classifies the borrower with three predictor variables such as age, daily income and daily payment of the borrower. Results of the study 22 out of 25 borrowers are correctly classified.

## 3. Basic Concepts and Methodology

### 3.1 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is widely used multivariate statistical method for analysis of data with categorical outcome variables. LDA is an appropriate for the development of linear classification models, i.e. models associated with linear boundaries between the groups. LDA has been developed for normally distributed explanatory variables. It is therefore reasonable to expect LDA to give better results in the case when the normality assumptions are fulfilled. (Pohar,2004). It can be applied in positioning and product management. Some studies use Linear Discriminant Analysis in bankrupt predictions, face recognition, credit scoring and the like.

#### 3.1.1 Linear Discriminant Analysis for Two Classes

Consider a set of observations  $x$  (also called features, attributes, variables or measurements) for each sample of an object or event with known class  $y$ . These set of samples are called the training set. The classification problem is then to find a good predictor for the class  $y$  of any sample of the same distribution (not necessarily from the training set) given only an observation  $x$ .

Linear Discriminant Analysis approaches the problem by assuming that the conditional probability density functions  $p(x|y=0)$  and  $p(x|y=1)$  are both normally distributed with mean and covariance parameters

$\mu_0, \sum_{y=0}$  and  $\mu_1, \sum_{y=1}$ , respectively. Under this assumption, the Bayes optimal solution is to predict points as being from the second class if the log of the likelihood ratios is below some threshold  $T$ , so that;

$$(x_i - \mu_0)^T \sum_{y=0}^{-1} (x - \mu_0) + \ln \left| \sum_{y=0} \right| - (x_i - \mu_1)^T \sum_{y=1}^{-1} (x - \mu_1) - \ln \left| \sum_{y=1} \right| < T$$

Without any further assumptions, the resulting classifier is referred to as QDA (quadratic discriminant analysis). LDA also makes the simplifying homoscedasticity assumption (i.e. that the class covariances are identical, so  $\sum_{y=0} = \sum_{y=1} = \sum$ ) and that the covariances have full rank. In this case, several terms cancel and the above decision criterion becomes a threshold on the dot product  $w \cdot x > c$  for some threshold constant  $c$ , where\

$$w = \sum^{-1} (\mu_1 - \mu_0) \quad (3.2)$$

This means that the criterion of an input  $x$  being in a class  $y$  is purely a function of this linear combination of the known observations.

#### 3.1.2 Linear Discriminant Analysis Formula

Using classification criterion to minimize total error of classification (TEC), we tend to make the proportion of object that it misclassifies as small as possible. TEC is the performance rule in the ‘long run’ on random samples of objects. Thus, TEC should be thought as the probability that the rule under consideration will misclassify an object. The classification rule is to assign an object to the group with highest conditional probability. This is called Bayes Rule. This rule also minimizes the TEC. If there are  $g$  groups, the Bayes’ rule is to assign the object to group  $i$  where  $P(i|x) > P(j|x)$  for  $\forall j \neq i$ . We want to know the probability  $P(i|x)$  if an object belongs to group  $i$ , from the given set of measurement  $x$ . In practice, however, the quantity  $P(i|x)$  is difficult to obtain. What we can get is  $P(x|i)$ . This is the probability of getting a particular set of measurement  $x$ , given that the object comes from group  $i$ . For example, after we know that the soap is good or bad, then we can measure the object (weight, smell, color etc.). What we want to know is, to determine the group of the soap (good or bad) based on the measurement only.

Fortunately, there is a relationship between the two conditional probabilities known as Bayes Theorem:

$$P(i|x) = \frac{P(x|i) \cdot P(i)}{\sum_{\forall j} P(x/j) \cdot P(j)} \quad (3.3)$$

Prior probability  $P(i)$  is probability about the group  $i$  known without making any measurement. In practice, we can assume the prior probability is equal for all groups or based on the number of sample in each group.

In practice, however, to use the Bayes rule the directly  $P(x|i)$  is impractical because to obtain  $P(x|i)$ , it needs more data to get the relative frequencies of each group for each measurement. It is more practical to assume the

distribution and get the probability theoretically. If we assume that each group has multivariate Normal distribution and all groups have the same covariance matrix, we get what is called Linear Discriminant Analysis formula:

$$f_i = \mu_i C^{-1} x_k^T - \mu_i C^{-1} \mu_i^T + \ln(p_i) \quad (3.4)$$

Assign object  $k$  to group  $I$  that has maximum  $f_i$ .

The second term  $(\mu_i C^{-1} \mu_i^T)$  is actually Mahalanobis distance, which is, distance to measure dissimilarity between several groups.

### 3.1.3 Fisher's Discriminant Function

To set the stage for developing significance test for the discriminant functions, let us assume that we have two random samples of observations on set of continuous variables from two multivariate normal populations, with assumed common covariance matrices whose centroids may differ.

We assumed that an estimate of the common population covariance matrix can be obtained by pooling the two samples covariance matrix  $C_w$  and have noted that it is proportional to  $W$ , the pooled within groups sum of squares and cross product (SSCP) matrix. In two group discriminant analysis, we seek a set of weights, we expressed by vector:

$$k = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$$

From which the following linear composite  $f$  - called the discriminant axis-can be obtained

$$f = Xdk = \begin{bmatrix} X_{d11} & X_{d12} & \cdots & X_{d1n} \\ X_{d21} & X_{d22} & \cdots & X_{d2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{dm1} & X_{dm2} & \cdots & X_{dmn} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

Where  $X_d$  is the  $m \times n$  matrix of mean-corrected scores on the original predictors. Fisher's linear discriminant function is defined as:

$$f_i = X'_{di} k = x_{di1} k + x_{di2} k + \dots + x_{din} k_n \quad (3.6)$$

where  $i$  = the students from each group  
 $X$  = set of all predictor variables  
 $k$  = discriminant weight

That is, each  $f_i$  denotes the score of the  $i$ -th person on the discriminant axis  $f$ . The mean of the linear composite  $f$  is found by substituting the centroid of  $X_d$  in the discriminant function:

$$F = \bar{x}'_{di} k = (0, 0, 0, \dots, 0) \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = 0 \quad (3.7)$$

That is,  $f = 0$  since the matrix  $X_d$  is already expressed in mean-corrected form.

## 3.2 Methodology

### 3.2.1 To develop discriminant rules based on the pre-determined factors of Computer Games Addiction.

This study used the Young Diagnostic Test (eight-item questionnaire) to use a pathological internet use which criteria for pathological gambling, (see Appendix E.). The study will conduct a survey in the form of questionnaire to the students who are playing computer games and a student in Mindanao University of Science and Technology (MUST). The said items will be used in the survey questionnaire. The study used Likert Scale to answer the survey questionnaire, 1- Strongly Disagree 2-Disagree 3- Neither 4- Agree 5- Strongly Agree.

Based on Poon (2012), the researcher categorized the player as addicted if the player answers 4-5 points/yes in each item and 1-3 points/yes answers equaled social playing and non-addicted. In this study, if the final score of each student ranges from 0 to 30, it will categorize as non-addicted (Group 1). If the final score of each student will ranges from 31 to 50, then it will categorize as addicted (Group 2).

### 3.2.2 To classify student as addicted or non-addicted to computer games using Linear Discriminant Analysis.

The researcher used Matlab R2009a in order to run the data gathered from survey questionnaire.

Linear discriminant analysis can be used to determine which variable discriminates between two or more classes, and to derive a classification model for predicting the group membership of new observations (Worth and Cronin, 2003). For each of the groups, LDA assumes the explanatory variables to be normally distributed with equal covariance matrices. The simplest LDA has two groups. To discriminate between them, a linear discriminant function that passes through the centroids of the two groups can be used.

### 3.2.3 To evaluate the performance of the discriminant function.

The study will test the significance of the discriminant function based on the Fisher Discriminant Function. This will provide us the total classification of the model. In this section, the Appendix D will show the misclassification occur in each group.

## 4. Results

### Preliminary Analysis

The data used in this study obtained through survey which shown in Appendix A. These data are used to classify

student's engagement. The following table shows the predicted variables and the names of the variables given to this data. For instance, X<sub>2</sub> is for variable or question "I feel preoccupied with playing computer games (think about previous gaming activity or anticipate next gaming sessions)?" and others will follow for X<sub>2</sub> to X<sub>9</sub>.

**Table 4.1: Variable Information**

Predictor variables	Variables Name
Age	X <sub>1</sub>
I feel preoccupied with playing computer games (think about previous gaming activity or anticipate next gaming sessions)	X <sub>2</sub>
I feel the need to play computer games with increasing amounts of time order to achieve satisfaction or for it to be enjoyable	X <sub>3</sub>
I repeatedly made unsuccessful efforts to control, cut back, or stop gaming activities	X <sub>4</sub>
I feel restless, moody, depressed, or irritable when attempting to cut down or stop gaming usage	X <sub>5</sub>
I play longer than originally intended	X <sub>6</sub>
I jeopardized or risked the loss of a significant relationship, educational, or career opportunity because of your gaming use	X <sub>7</sub>
I lied to my family members, a therapist, or others to conceal how much you play computer games	X <sub>8</sub>
I play computer games as a way of escaping from problems or relieving feelings of helplessness, guilt, anxiety, or depression.	X <sub>9</sub>
Average Hours of playing in a day	X <sub>10</sub>

**Developing discriminant rules based on the pre-determined factors of Computer Games Addiction.**

The table in Appendix B shows the raw data of the students obtained from survey questionnaire. In this table, students were grouped into two as addicted or non-addicted. The 12<sup>th</sup> column shows the groupings of each student. A non-addicted (Group 1) is denoted by 0 and addicted (Group 2) is denoted by 1. The groupings are based after all the questions have been answered. The researcher added the numbers for each response from X<sub>1</sub> to X<sub>9</sub> to obtain a final score which will be found in the 11<sup>th</sup> column. If the final score of each student ranges from 0 to 30, it will categorize as non-addicted (Group 1). If the final score of each student will ranges from 31 to 50, then it will categorize as addicted (Group 2).

The X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, ..., X<sub>9</sub> from the Table 3.1 are the predictor variables of the criterion variable (Y). Using these variables, the mean of each group is:

$$\mu_1 = [2.5385 \quad 3.1231 \quad 3.0769 \quad 3.0308 \quad 2.6769 \quad 2.8462 \quad 2.1846 \quad 2.0923 \quad 3.2462 \quad 1.1692]$$

$$\mu_2 = [3.0286 \quad 3.8000 \quad 4.2571 \quad 3.8857 \quad 3.7429 \quad 3.8857 \quad 3.2857 \quad 3.1429 \quad 4.4286 \quad 1.6286]$$

And the mean for the total number of students is:

$$\mu = [2.7100 \quad 3.3600 \quad 3.4900 \quad 3.3300 \quad 3.0500 \quad 3.2100 \quad 2.5700 \quad 2.4600 \quad 3.6600 \quad 1.3300]$$

The Equ. 4.1 is used to obtain the total sample deviation given by:

$$X_d = X - \mu \tag{4.1}$$

Where  $\mu$  is the over sample mean,  $X_d$  is the deviations from overall-sample mean and  $X$  is the predictor variable. The table in Appendix C summaries all the deviations from the overall centroid. As we can see in the table, group 1 has negative deviation compared to group 2. Group 1's mean deviations are -0.1715, -0.2369, -0.4131, -0.2992, -0.3731, -0.3638, -0.3854, -0.3677, -0.4138 and -0.1608, respectively while the Group 2's mean deviations are 0.3186, 0.4400, 0.7671, 0.5557, 0.6929, 0.6757, 0.7157, 0.6829, 0.7686 and 0.2986, respectively. As can be observed, the first group consists of negative mean deviations while the second group consists of positive mean deviations. This implies that the students from each group are separated from each other. The problem is to find a discriminant function to rotate the variables in order to maximize the distance between groups and to minimize the distance within groups. Then, we can draw a line connecting the sub-groups centroids. Through this line, we can project each of the points and assign each point to the closer of the two centroids. The discriminant function is given by Equation 3.6. To find the discriminant function, the Equ. 4.2 used to find the pooled-within group covariance.

$$C_w = \frac{W}{\sum_{i=1}^2 (n_i = 1)} \tag{4.2}$$

where  $W$  is the within-group sum of squares and cross-product. That is,  $W = P'P$ ,  $P$  is the matrix of within-group deviations. Thus,

$$C_w = \begin{bmatrix} 0.7054 & 0.0091 & -0.0505 & -0.0302 & 0.0160 & -0.0663 & 0.1352 & -0.1569 & -0.1331 & 0.0454 \\ 0.0091 & 0.7410 & 0.2876 & 0.0505 & 0.1203 & 0.0860 & 0.0053 & -0.1402 & 0.0411 & 0.0821 \\ -0.0505 & 0.2876 & 0.8704 & 0.1008 & 0.0401 & 0.0490 & -0.0561 & -0.1811 & 0.1522 & 0.0051 \\ -0.0302 & 0.0505 & 0.1008 & 0.9539 & 0.1288 & -0.1036 & -0.0125 & -0.0267 & 0.1247 & 0.0018 \\ 0.0160 & 0.1203 & 0.0401 & 0.1288 & 0.9684 & 0.0892 & 0.0658 & 0.1758 & -0.0099 & 0.0226 \\ -0.0663 & 0.0860 & 0.0490 & -0.1036 & 0.0892 & 0.6531 & 0.0611 & -0.0256 & 0.1447 & 0.0837 \\ 0.1352 & 0.0053 & -0.0561 & -0.0125 & 0.0658 & 0.0611 & 0.8258 & 0.0864 & 0.0384 & -0.0951 \\ -0.1569 & -0.1402 & -0.1811 & -0.0267 & 0.1758 & -0.0256 & 0.0864 & 1.1197 & 0.0039 & 0.0188 \\ -0.1331 & 0.0411 & 0.1522 & 0.1247 & -0.0099 & 0.1447 & 0.0384 & 0.0039 & 0.9452 & -0.0218 \\ 0.0454 & 0.0821 & 0.0051 & 0.0018 & 0.0226 & 0.0837 & -0.0951 & 0.0188 & -0.0218 & 0.3195 \end{bmatrix}$$

And the inverse within-group covariance is

$$C_w^{-1} = \begin{bmatrix} 1.6460 & 0.0251 & 0.0815 & 0.0606 & -0.0823 & 0.2380 & -0.3614 & 0.3009 & 0.1762 & -0.4117 \\ 0.0251 & 1.6669 & -0.5048 & -0.0164 & -0.1908 & -0.0901 & -0.0953 & 0.1725 & 0.0196 & -0.4239 \\ 0.0815 & -0.5048 & 1.4045 & -0.0882 & -0.0246 & -0.0167 & 0.0902 & 0.1686 & -0.1806 & 0.1071 \\ 0.0606 & -0.0164 & -0.0882 & 1.1386 & -0.1811 & 0.2729 & -0.0123 & 0.0569 & -0.1722 & -0.0869 \\ -0.0823 & -0.1908 & -0.0246 & -0.1811 & 1.1440 & -0.1913 & -0.0421 & -0.2253 & 0.0694 & 0.0369 \\ 0.2380 & -0.0901 & -0.0167 & 0.2729 & -0.1913 & 1.7963 & -0.2177 & 0.1241 & -0.2774 & -0.5602 \\ -0.3614 & -0.0953 & 0.0902 & -0.0123 & -0.0421 & -0.2177 & 1.3796 & -0.1620 & -0.0694 & 0.5498 \\ 0.3009 & 0.1725 & 0.1686 & 0.0569 & -0.2253 & 0.1241 & -0.1620 & 1.0399 & -0.0239 & -0.2177 \\ 0.1762 & 0.0196 & -0.1806 & -0.1722 & 0.0694 & -0.2774 & -0.0694 & -0.0239 & 1.1822 & 0.1030 \\ -0.4117 & -0.4239 & 0.1071 & -0.0869 & 0.0369 & -0.5602 & 0.5498 & -0.2177 & 0.1030 & 3.6239 \end{bmatrix}$$

Now that  $C_w^{-1}$  is already obtained, we are now to find the discriminant weights to formulate the discriminant function. Thus,

$$k = C_w^{-1}d \tag{4.3}$$

where  $d = X_{d1} - X_{d2}$  is the differences vector of centroids and,  $X_{d1}$  and  $X_{d2}$  are the mean deviations of each group.

Thus,

$$d = [-0.4901 \quad -0.6769 \quad -1.1802 \quad -0.8549 \quad -1.0659 \quad -1.0396 \quad -1.1011 \quad -1.0505 \quad -1.1824 \quad -0.4593]$$

obtaining the mean vectors of group 1 and 2 in the discriminant function. That is,

$$f_1 = k' X_{d1}$$

Therefore,

$$k = C_w^{-1} d = \begin{bmatrix} -1.0689 \\ -0.1386 \\ -1.3490 \\ -0.7814 \\ -0.4832 \\ -1.2380 \\ -1.1025 \\ -1.1866 \\ -0.8684 \\ -1.1834 \end{bmatrix}$$

Since,  $k$  and  $X_d$  are already obtained we can compute for the discriminant scores of each group. This can be done by

$$= [-1.0689 - 0.1386 - 1.3490 - 0.7814 - 0.4832 - 1.2380 - 1.1025 - 1.1866 - 0.8684 - 1.1834] \begin{bmatrix} -0.1715 \\ -0.2369 \\ -0.4131 \\ -0.2992 \\ -0.3731 \\ -0.3638 \\ -0.3854 \\ -0.3677 \\ -0.4138 \\ 0.1608 \end{bmatrix}$$

$$= 3.0487$$

$$f_2 = k' X_{d2}$$

$$= [-1.0689 - 0.1386 - 1.3490 - 0.7814 - 0.4832 - 1.2380 - 1.1025 - 1.1866 - 0.8684 - 1.1834] \begin{bmatrix} 0.3186 \\ 0.4400 \\ 0.7671 \\ 0.5557 \\ 0.6929 \\ 0.6757 \\ 0.7157 \\ 0.6829 \\ 0.7686 \\ 0.2986 \end{bmatrix}$$

$$= -5.6621$$

We can now compute for the discriminant scores of each student. The results are summarized in Appendix D.

**Classifying student as addicted or non-addicted to computer games using Linear Discriminant Analysis**

Computing the cutoff point  $f_c$  the halfway between the mean discriminant scores of groups 1 and 2:

$$f_c = \frac{(f_1 + f_2)}{2} = \frac{3.0487 + -5.6621}{2} = -1.3066$$

If the student discriminant score was greater than the cut-off point, then that student was labeled as non-addicted, otherwise labeled as addicted. In Appendix D, the first student discriminant score is  $-1.1117$  and the cutoff point is  $-1.3066$ . That is,  $-1.1117 > -1.3066$ , implies that the first student is classified to group 1 as no-addicted. Therefore, in order for a student to be labeled as belonging to the group that has low risk in addiction the discriminant score of a student must be at least  $-1.3066$  and above.

#### 4.4 Evaluating the performance of the discriminant function

Results showed in the said Appendix D that in the non-addicted group, 61 out of 63 or 96.82% of the student are correctly classified and in the addicted group, 35 out of 37 or 94.59% of the student are correctly classified. Moreover, there is 4.29% of average misclassification probability which implies that Linear Discriminant Analysis perform better in classifying behavioral addiction.

### 5. Conclusions and Recommendations

Linear discriminant analysis (LDA) is widely used multivariate statistical method for analysis of data with categorical outcome variables. It is one of the methods used in statistics to find a linear combination of features which characterizes or separates two or more classes of objects or events and to derive a classification model for predicting the group membership of new observations.

The researcher gathered data from Mindanao University of Science and Technology (MUST) using survey questionnaire. The survey questionnaire includes questions regarding Computer Game Engagement adopted from the Young's Diagnostic Test for Internet Addiction Test and used Likert Scale in order to answer to the questions from surveys. By using Linear Discriminant Analysis, the researcher was able to classify students as addicted or non-addicted to computer games. The function is given by equation and through this the researcher was able to obtain the discriminant score of each group. Lastly, the researcher was able to classify students as addicted or non-addicted and found out that in non-addicted group, 61 out of 63 or 96.82% of the student are correctly classified and in the addicted group, 35 out of 37 or 94.59% of the student are correctly classified. Moreover, the average misclassification probability is 4.29% which implies that the Linear Discriminant Analysis performs better in classifying behavioral addiction.

Students tend to play while multi-tasking, so they don't spent time on one activity. The result shows that the students in MUST are not fun in computer games. Computer gaming seems to be a recreational past time in Mindanao University of Science and Technology and does not seem to have detrimental effects on the student body.

Research already indicates that computer games can be destructive presence in a game player's life, but can also have a positive impact. Gaming allows players to communicate with others they may not ordinarily meet, gives self-confidence, and develops skills that may be used in real life. As demonstrated above, there is a small portion of computer game users who are at risk for computer game addiction.

For further improvements of this study, the researcher wants to recommend to classify players into more than two groups, use other methods in classifying players and compare it to Linear Discriminant Analysis and categorize players as to computer game addiction level.

### References

- [1] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. Washington, DC: American Psychiatric Association, 2000.
- [2] Ahn, J. and Randall, G. Computer Game Addiction. Clemson University. pp.1-3
- [3] Brenner, V. (1996). An initial report on the on-line assessment of Internet addiction: The first 30days of the Internet usage survey. <http://www.ccsnet.com/prep/pap/pap8b/638b012p.txt>
- [4] Cabangca, M. (2011). Credit Risk Analysis via Linear Discriminant Analysis.
- [5] Casio, R. (2011). Development of classification models for predicting success of electronics engineering graduates of MUST on licensure examination.
- [6] Griffiths, M. and Widyanto, L. (2006). Internet Addiction': A Critical Review. International Gaming Research Unit, Psychology Division, Department of Social Sciences, Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, UK
- [7] Langley, Alex. Correlates of Video Game Addiction. Master of Science (Psychology), December 2010, 51 pages, 4 tables, references, 37 titles.
- [8] Pohar, M. et. al (2004). Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. Metodološki Zvezki, Vol. 1, pp. 143-161.
- [9] Poon, Amy W., "Computer Game Addiction and Emotional Dependence". Senior Theses, Trinity College, Hartford, CT 2012. Trinity College Digital Repository, <http://digitalpository.trincoll.edu/theses/253>
- [10] Shaw, M., Black, D. W. (2008). Internet Addiction. CNS Drugs, 22(5), 353-365.
- [11] Sahin, C. (2011) "An Analysis of Internet Addiction Levels of Individual According to Various Variables". TOJET: The Turkish Online Journal of Educational Technology – October 2011, volume 10 Issue 4
- [12] Supaket, P. et al., (2008). "Factors Affecting Computer Game Addiction and Mental Health of Male Adolescents in Mueang District, Si Sa Ket Province". Journal of Public Health, Vol. 38 No. 3
- [13] Tuisima, S. et al. "Classification of Computer Game Addiction Level in Students in Secondary Education (M.1-3) using Neural Networks". 2012 2nd International Conference on Management and Artificial Intelligence IPEDR Vol.35 (2012) © (2012) IACSIT Press, Singapore
- [14] Worth, A.P. and Cronin, M.T.D. (2003): The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. Theochem, 622, 97-111.
- [15] Yeoh, W., Computer addiction ends in tragedy. Retrieved from <http://www.thestar.com.my/News/Nation/2013/11/06/Computer-addiction-ends-in-tragedy-23-year-old-found-slumped-over-the-keyboard-in-front-of-his-computer.aspx/>

- [16] Young, K. (1996a). Internet Addiction: The Emergence of a New Clinical Disorder. *Cyber Psychology and Behavior*, 3, 237–244.
- [17] Young, K. (1998). Assessment of Internet Addiction. The Center for Internet Addiction Recovery. pp 4-5. [www.netaddiction.com](http://www.netaddiction.com)

### Author Profile



**Kennet G. Cuarteros** graduated his Master of Science in Applied Mathematics in Mindanao University of Science and Technology and also his Bachelor of Science in Applied Mathematics degree in the same university. Currently he is a full-time faculty and the coordinator of the Department of Mathematical Sciences in University of Science and Technology of Southern Philippines.



**Rose May L. Puerte** graduated with Bachelor of Science in Applied Mathematics degree in Mindanao University of Science and Technology last March 2014.