DOI : 10.18843/rwjasc/v10i3/01
DOI URL : http://dx.doi.org/10.18843/rwjasc/v10i3/01

Classification and Prediction of Student's GPA using Fisher Linear Discriminant (FLD) Function

R. Gunawan Santosa,

Department of Informatics Duta Wacana Christian University, Indonesia

Antonius Rachmat Chrismanto,

Department of Informatics Duta Wacana Christian University, Indonesia

Erick Kurniawan,

Department of Information System Duta Wacana Christian University, Indonesia

ABSTRACT

Classification and prediction are some of the capabilities of Data Mining. This study will implement a classification model using the Fisher Linear Discriminant (FLD) function. After the classification model is obtained, the model is used to predict the Grade Point Average category (GPA-1st). The FLD classification models used are 9 models derived from cumulative student data from 2008 to 2016 academic year. In the FLD model, GPA-1st is used as the dependent variable, while the factors of high school location, high school status, high school type, and English proficiency level are used as independent variables. These models are used to predict the GPA-1st category for students in 2017. Crosstab tables are used to measure the accuracy of the classification model and accuracy of the prediction model. As the result, the accuracy average of the 9 classification models in students' GPA-1st is 68.67%. While the accuracy average of predictions using 9 models is 58.28%.

Keywords: classification, prediction, crosstab table, Fisher Linear Discriminant (FLD).

INTRODUCTION:

Data Mining:
According to (Larose & Larose, 2014) Central Connecticut State University's Director of Data Mining, the following are some of the most common tasks and objectives in Data Mining topics.

Description:
At this point, researchers and analysts only try to find ways to describe the patterns and trends in the data.

Estimation:
Estimates are usually related to mathematical models based on the general population parameters. Usually these parameters are not known so we can only guess or estimate the value of these parameters with samples taken from the population.

Prediction:
Predictions are like classifications and estimates, except for predictions, the results are in the future.

Classification:
In classification, there are target variables which include categorical data, such as income categories, which can be divided into three categories, namely: high income, middle income, and low income.

Clustering:
Clustering refers to grouping data or records into classes similar. Clusters are the process of grouping data that are like each other and are different from other cluster groups. Clustering differs from classification in the absence of a target variable for grouping. Clustering algorithms usually look for segmentation of all data sets into relatively homogeneous subgroups or clusters, where the similarity of data in a cluster is

maximized and its similarities outside the cluster are minimized.

Association:

The association in Data Mining is to find out which attributes "go together." Most prevalent in the business world, where it is known as affinity analysis or Market Basket Analysis, the purpose of the association seeks to uncover rules to measure the relationship between two or more attributes. The association rule is a form of "if antecedent, then consequent," along with a measure of support (support) and confidence (beliefs) related to a rule (Han, Kamber, & Pei, 2011).

Each of these approaches can be used to analyze quantitatively a large set of data to get hidden meaning (pattern) in a system. Data Mining is usually an exploration process, but it can also be used for confirmation investigations (Berson, Smith, & Thearling, 2011).

EDUCATIONAL DATA MINING:

In its development, Data Mining can also be applied in the field of education. With this application, a research called EDM (Educational Data Mining) developed. Educational Data Mining (EDM) is an emerging field to explore data in the context of education by applying various Data Mining (DM) techniques / tools. This will provide intrinsic knowledge in the teaching and learning process for more effective educational planning. Nowadays there have been many research studies in education that use Data Mining.

In a society that is supported by increasingly fast Information Technology (IT), heterogeneous data mining is an important issue, therefore (Jindal & Borah, 2013) wrote about a journey of research and practice in the field of EDM from 1998 to 2012. This paper focuses on Offline, Online research trends and uncertain but useful data sources in the field of education in the future. This research can be further enhanced to design Knowledge Findings based on Decision Support Systems (Knowledge Discovery based Decision Support Systems) that will be able to provide the right decisions for scientific and technological research based on community demand. As a continuation of this they will solve two problems, namely building a real model that relates to the application of tools and techniques specifically and the second to make predictive models using data in an ongoing manner (incremental data).

Ayala A. P. conducted a survey on research related to EDM. The two objectives of the survey are, the first is to preserve and enhance the latest EDM studies that advance the education process; while the second is managing, analyzing, and discussing reviews based on the results and benefits obtained in the Data Mining (DM) approach. Because of the 240 analysis of EDM work, the EDM work profile was compiled to illustrate 222 EDM approaches and 18 methods or algorithms used. One important finding is: most EDM approaches are based on a basic set of objects compiled by three types of educational systems namely discipline, tasks and methods (algorithms). The conclusion is a survey of EDM that relates to the strengths, weaknesses, opportunities, and threats that are expected to be faced by research in the future (Ayala, 2014).

Romero and Ventura (Romero & Ventura, 2013) in their research presented several topics related to EDM, namely:

- Data analysis and visualization
- Provide feedback to support instructors
- Recommendations for students
- Predict student performance
- Modeling students
- Detect unwanted student behavior
- Grouping of students
- Social network analysis
- Develop a concept map
- Build courseware
- Planning and scheduling

It is known that the evaluation of the success of learning in college is measured by Grade Point Average (GPA). The beginning of lectures in college is an important matter because it is a transition period from high school education to scholarship education. From the observation, it turned out that there were several students who received high Achievement Indexes, but there were students who had very low Achievement Indexes (Romero & Ventura, 2013), (Santosa & Setiadi, 2015).

BACKGROUND OF STUDY:

Likewise, what happened at Duta Wacana Christian University, especially in the Information Technology Faculty, it turned out that in the first semester there were several students who had a low GPA-1st. So, there are some students who have not registered in the following semesters and there are students who move to other faculties (Santosa & Setiadi, 2015). This creates problems that disrupt the college system. In order to reduces this problem, a research was conducted to create a model for grouping or classifying new students accepted at FTI UKDW with the DLF function. Then the model can be used to predict the GPA-1st category. In this case, we take two EDM topics according to (Romero & Ventura, 2013) namely grouping of students and predict student performance. Performance prediction in this case is to categorize students' GPA-1st in two categories, namely students with low GPA-1st or students with high GPA-1st using FLD functions.
In the research that had been done (Santosa & Chrismanto, 2016), it was obtained the prediction results of the GPA-1st category with the Logistic Regression model for students of 2015 who through academic achievement track has an accuracy average of 67.87%. While the result of the accuracy average of predictions for the GPA-1st category with Logistic Regression for students in the 2016 through academic achievement was 67.8013%. Both research results are not much different (Santosa & Chrismanto, 2017).

LITERATURE REVIEW:

According to (Rencher, 2002) there are two main objectives in group separation, such as:

1. Description of group separation, where the linear function of a variable (discriminant function) is used to describe or explain the difference between two or more groups. The purpose of this descriptive analysis includes identifying the contribution of several variables to group separation and is usually referred to as classification.
2. Prediction is the allocation of an observation to a group, usually using the classification function. This function is used to assign an individual to a group. Usually the measured value of the individual or object is calculated by the classifier function so that a group can be found which is the most likely individual presence in the group.

One of the classifying functions is Fisher Linear Discriminant function. Some studies that use the FLD function, for example as below.

- Alhaddad et.al diagnosed an Electroencephalogram (EEG) -based Autism using Fisher Linear Discriminant (FLD) analysis. The result is that the FFT feature signal has an accuracy level higher than 88.14% and a standard deviation of 0.0404 lower than the original feature (Alhaddad, et al., June 2012).
- Alexandre et.al conducted research on the classification of speech or music using Fisher Linear Discriminant analysis. It turns out that the results given show that the behavior of the FLD classification algorithm is better than that of the closest K-neighbor algorithm. This will also show that it is possible to get very good results in terms of probability with the smallest errors. And by using only one feature extracted from audio signals, it is possible to reduce the complexity of the application system in real-time (Alexandre, Cuadra, Gil-Pita, & Rosa, 2006).
- Dumitra evaluate the company's performance through Discriminant analysis of 20 companies traded on the Bucharest Stock Exchange (BVB). Because these companies are similar in terms of business profiles (ie: manufacturing industry), ten financial indicators related to stock values are selected (PRICE, BETA, ALPHA, etc.) and book values (Debt / Equity, ROA and ROE) to assess and classify companies as "good" or "bad". For sustainable characterization, the average value of financial indicators is estimated between the first quarter of 2005 and the third quarter of 2013. Initial groupings are made according to return on assets (ROA) and divide the sample into 10 "good companies" and 10 "bad companies". The result is that discriminant analysis correctly validates the classification of companies with ROA criteria in 90% of cases (18 of 20 companies). In addition, the analysis also shows that ROA is the first important in evaluating company performance as suggested by test statistics (Dumitra & Tudor, November 6th - 7th ,2014).

This research is essentially including Educational Data Mining. Data mining has roots in machine learning, artificial intelligence, computer science, and statistics (Dunham, 2002). Different methods of data mining approaches are grouping, classifying, and building association rules. Linear Fisher's Discriminant function model is included in the topic of multivariate statistics (Johnson & Wichern, 2008).
Some studies which used FLD model in prediction studies, such as:

- The Discriminant function can be used to get a sharp classification to distinguish between students who are accepted and those who are not accepted in the department of Industrial Chemistry for the class of 2009/2010. The results of the analysis show that the average value of the University Matriculation Examination (UME) by accepted candidates is higher than the average value of candidates who are not accepted (Okoli, 2013).
- Discriminant models have also been used to predict student financial loans, whether they will be paid off or will fail. The result is that the accuracy of the system is 77% (Muthii, 2015).
- Ogum used the multivariate analysis method to analyze the score of candidates admitted in the medical department of the University of Nigeria in the academic year 1975/76 and formed a discriminant function that could discriminate between students who were accepted and not accepted (Ogum, 2002).
- Research on classification has been carried out (Santosa, 2008) with FTI student data from 2001 to 2003 and it turns out that the Fisher Linear Discriminant (FLD) function model can classify student with an average accuracy of 62%, but in that study the results of the DLF function have not been used for prediction.

In this section we will discuss a few bases derived from mathematical-statistical theory that will underlie the classification function of Fisher Linear Discriminant (Johnson & Wichern, 2008). Vector notation $\mathbf{X}^T = [X_1, X_2, \ldots X_p]$ is used as multivariate observation consisting of p factors or components. Whereas π_1 is a symbol of population 1 and π_2 is a symbol of population 2.
Fisher's idea in determining the Linear Discriminant function for this classification is to transform (change) multivariate observation **X** into univariate observation y so that y can be classified as part of the population π_1 or part of the population π_2.
Fisher proposed forming a linear combination of X to form Y, because this is a simple and easy function, so this function is known as Linear Fisher's Discriminant. First defined:
$\boldsymbol{\mu_1} = E(\mathbf{X} \mid \pi_1)$ = multivariate observation expectation value π_1
$\boldsymbol{\mu_2} = E(\mathbf{X} \mid \pi_2)$ = multivariate observation expectation value π_2
and $\boldsymbol{\Sigma} = E(\mathbf{X}-\boldsymbol{\mu}_1)(\mathbf{X}-\boldsymbol{\mu}_2)^T$. In Multivariate Statistics, $\boldsymbol{\mu}$ is the measure of center while $\boldsymbol{\Sigma}$ is the measure of spread. Then a linear combination of $Y = \mathbf{l}^T\mathbf{x}$ is defined for a vector **l** and Y which is an univariate form.
$\mu_{1Y} = E(\mathbf{l}^T\mathbf{X} \mid \pi_1) = \mathbf{l}^T\boldsymbol{\mu}_1$ = multivariate observation expectation value π_1 in Y.
$\mu_{2Y} = E(\mathbf{l}^T\mathbf{X} \mid \pi_2) = \mathbf{l}^T\boldsymbol{\mu}_2$ = multivariate observation expectation value π_2 in Y. Whereas variance Y is symbolized by $\sigma_Y^2 = \mathrm{Var}(\mathbf{l}^T\mathbf{X}) = \mathbf{l}^T\mathrm{COV}(\mathbf{X})\,\mathbf{l} = \mathbf{l}^T\boldsymbol{\Sigma}\,\mathbf{l}$.
The linear combination of the best **L** values can be found from comparison:

$$\frac{(squared\ \ distance\ \ between\ \ mean\ \ Y\ \ )}{(variance\ \ Y)} = \frac{(\mu_{1Y}-\mu_{2Y})^2}{\sigma_Y^2}$$

$$= \frac{\left(\ell^T\boldsymbol{\mu}_1-\ell^T\boldsymbol{\mu}_2\right)^2}{\ell^T\Sigma\,\ell} = \frac{\left(\ell^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right)^2}{\ell^T\Sigma\,\ell} = \frac{\left(\ell^T\boldsymbol{\delta}\right)^2}{\ell^T\Sigma\,\ell}.$$

Our goal is to look for a linear combination coefficient $\mathbf{l}^T = [\,l_1\, l_2\, l_3\ l_4 \ldots l_p\,]$ the best is $\mathbf{l}^T = [\,l_1\, l_2\, l_3\, l_4 \ldots l_p]$ which maximizes $\frac{\left(\ell^T\boldsymbol{\delta}\right)^2}{\ell^T\Sigma\,\ell}$. This linear combination coefficient $\mathbf{l}^T = [\,l_1\, l_2\, l_3\, l_4 \ldots l_p]$ is often called the Linear Fisher Classification Coefficient.
To get the value of the Fisher Linear Classification Coefficient, some supporting theorems are needed (Johnson & Wichern, 2008) below:

Theorem 2.1 (Cauchy-Schwarz Inequality):
Given $\bar{b}$ and $\bar{d}$ are 2 vectors that have size px1, then

$$\left(\bar{b}^T\bar{d}\right)^2 \le \left(\bar{b}^T\bar{b}\right)\left(\bar{d}^T\bar{d}\right) \qquad (1)$$

and the form above becomes an equation if and only if $\bar{b} = c\bar{d}$ or $\bar{d} = c\bar{b}$ for a constant c.
Theorem 2.2 (The generalized Cauchy-Schwarz inequality):
Given $\bar{b}$ and $\bar{d}$ are 2 vectors that have size px1 and **B** is a positive definite matrix having a size of pxp , then

$$\left(\bar{b}^T\bar{d}\right)^2 \le \left(\bar{b}^T\mathbf{B}\bar{b}\right)\left(\bar{d}^T\mathbf{B}^{-1}\bar{d}\right) \quad (2)$$

and the form above becomes an equation if and only if $\bar{b} = c\mathbf{B}^{-1}\bar{d}$ or $\bar{d} = c\mathbf{B}\bar{b}$ for a constant c.

Theorem 2.3 (Spectral Decomposition):
Given A is a symmetrical matrix that has a size of p x p and λ_i , i = 1,2,3, …p is eigenvalue of **A** as well $\mathbf{e}_i$ is a normal eigenvector from λ_I then $\mathbf{A} = \lambda_1\mathbf{e}_1\mathbf{e}_1^T + \lambda_2\mathbf{e}_2\mathbf{e}_2^T + \lambda_2\mathbf{e}_2\mathbf{e}_2^T + \ldots + \lambda_p\mathbf{e}_p\mathbf{e}_p^T$ or $\mathbf{A} = \sum_{i=1}^{p}\lambda_i\mathbf{e}_i\mathbf{e}_i^T$ due to the Spectral Decomposition Theorem: If A is a p x p symmetrical matrix and $\mathbf{A} = \sum_{i=1}^{p}\lambda_i\mathbf{e}_i\mathbf{e}_i^T$ the normal eigenvector becomes a matrix $\mathbf{P} = [\,\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{e}_3\ \ldots\ \mathbf{e}_p]$ then $\mathbf{A} = \sum_{i=1}^{p}\lambda_i\mathbf{e}_i\mathbf{e}_i^T = \mathbf{PBP}^T$ with $\mathbf{PP}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}$ and

$$\mathbf{B} = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ : & : & : & \ldots & : \\ 0 & 0 & 0 & \ldots & \lambda_p \end{bmatrix}$$ also $\lambda_i > 0$

Properties 1:

$$\mathbf{B}^{-1} = \mathbf{PB}^{-1}\mathbf{P}^T = \sum_{i=1}^{p}\frac{1}{\lambda_i}\mathbf{e}_i\mathbf{e}_i^T \quad (3)$$

Properties 2:

$\mathbf{B}^{1/2}$ is a diagonal matrix with a diagonal i-th element is $\sqrt{\lambda_i}$ and $\mathbf{A}^{1/2} = \sum_{i=1}^{p}\sqrt{\lambda_i}\mathbf{e}_i\mathbf{e}_i^T = \mathbf{PB}^{1/2}\mathbf{P}^T$ then $\left(\mathbf{A}^{1/2}\right)^{-1} = \sum_{i=1}^{p}\frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i\mathbf{e}_i^T = \mathbf{PB}^{-1/2}\mathbf{P}^T$ with $\mathbf{B}^{-1/2}$ is a diagonal matrix with a diagonal i-th element is $\frac{1}{\sqrt{\lambda_i}}$.
A linear combination coefficients $\mathbf{l}^T = [\,l_1\ l_2\ l_3\ l_4\ \ldots\ l_p]$ the best is $\mathbf{l}^T = [\,l_1\ l_2\ l_3\ \ l_4\ \ldots\ l_p\ ]$ which maximizes $\frac{\left(\ell^T\boldsymbol{\delta}\right)^2}{\ell^T\Sigma\ell}$, this will be achieved as in equation (4) :

$$\ell = \Sigma^{-1}\delta \text{ or } \ell = \Sigma^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right) \quad (4)$$

Proof:
This proof uses the above theorem (Johnson & Wichern, 2008). While (Hardle, 2015) use differentials on Calculus Multivariate to maximize this form. Given Σ is a definite positive matrix.
By applying the generalized Cauchy-Schwarz inequality as in equation (2): $\left(\bar{b}^T\bar{d}\right)^2 \le \left(\bar{b}^T\mathbf{B}\bar{b}\right)\left(\bar{d}^T\mathbf{B}^{-1}\bar{d}\right)$, divide the two parts with $\left(\bar{b}^T\mathbf{B}^{-1}\bar{b}\right) > 0$, so it will be obtained:

$$\frac{\left(\bar{b}^T\bar{d}\right)^2}{\bar{b}^T\mathbf{B}\bar{b}} \le \left(\bar{d}^T\mathbf{B}\bar{d}\right) \quad \text{o} \quad \underset{b}{maks}\frac{\left(\bar{b}^T\bar{d}\right)^2}{\bar{b}^T\mathbf{B}\bar{b}} = \left(\bar{d}^T\mathbf{B}\bar{d}\right) \quad (5)$$

The maximum value of the form will be achieved if taken $\bar{b} = c\mathbf{B}^{-1}\bar{d}$.
Take $\bar{b}^T = \ell^T$, $\bar{d} = \delta$, c=1 and $\mathbf{B} = \Sigma$ apply to equation (5), the results are obtained as shown in equation (4).
To get Fisher Linear Discriminant function can be obtained by the equation

$$Y = \mathbf{l}^T\mathbf{x} = \left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)\mathbf{S}_{pooled}^{-1}\mathbf{x} \quad \text{with}$$

$$\bar{\mathbf{x}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbf{x}_{1j}$$ (average vector of population 1)

$$\bar{\mathbf{x}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2}\mathbf{x}_{2j}$$ (average vector of population1)

$$\mathbf{S}_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}\left(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1\right)\left(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1\right)^T$$ (matrix of variance-covariance population 1)

$$\mathbf{S}_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_{22}}\left(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2\right)\left(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2\right)^T$$ (matrix of variance-covariance population 2)

and $$\mathbf{S}_{pooled} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{(n_1 + n_2 - 2)}$$

To get the allocation rules for Fisher Linear Discriminant function, a limit can be used:

$$m = \frac{centroid\_1 \;+\; centroid\_2}{2} \quad \text{if} \quad n_1 = n_2$$

$$m = \frac{n_2 * centroid\_1 \;+\; n_1 * centroid\_2}{n_1 + n_2} \quad \text{if} \quad n_1 \neq n_2$$

where

$$centroid\_1 = \left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)\mathbf{S}^{-1}_{pooled}\bar{\mathbf{x}}_1$$

$$centroid\_2 = \left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)\mathbf{S}^{-1}_{pooled}\bar{\mathbf{x}}_2$$

Suppose given an observation $\mathbf{x}_o$ then the y value is obtained by $y_o = \left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)\mathbf{S}^{-1}_{pooled}\mathbf{x}_o$

Allocate $\mathbf{x}_o$ in the population π_1 if $y_o < m$ and

Allocate $\mathbf{x}_o$ in the population π_2 if $y_o \geq m$

By using an existing application, the **l** coefficient in standard form can be obtained (Rencher, 2002) as in equation 6, namely:

$Y = \mathbf{l}^T\mathbf{x}$ or

$$Y = [\,l_1 \;\; l_2 \;\; l_3 \;\; l_4 \ldots l_p\,]\left[\frac{x_1 - \bar{x}_1}{s_1} \quad \frac{x_2 - \bar{x}_2}{s_2} \quad \ldots\ldots \quad \frac{x_p - \bar{x}_p}{s_p}\right]^T$$

or

$$Y = \ell_1\left(\frac{x_1 - \bar{x}_1}{s_1}\right) + \ell_2\left(\frac{x_2 - \bar{x}_2}{s_2}\right) + \ldots + \ell_p\left(\frac{x_p - \bar{x}_p}{s_p}\right) \qquad (6)$$

The form equation (6) can also be written (Hardle, 2015):

$Y = A_0 + A_1(x_1) + A_2(x_2) + A_3(x_3) + \ldots + A_p(x_p)$ (7)

While to calculate m can also use $E(Y) = \bar{Y}$, namely:

$m = E(Y) = A_0 + A_1E(x_1) + A_2E(x_2) + A_3E(x_3) + \ldots + A_pE(x_p)$ (8)

with E (xi) is the expected value or the average independent variable I.

In this research, a FLD function model was created to predict the GPA-1st of FTI UKDW student. In the FLD function model, the independent variables are high school status (x_1), high school location (x_2), high school type (x_3), level of English proficiency (x_4). While the dependent variable is the GPA-1st The Fisher Linear Discriminant (FLD) function model we built is shown in Figure 1

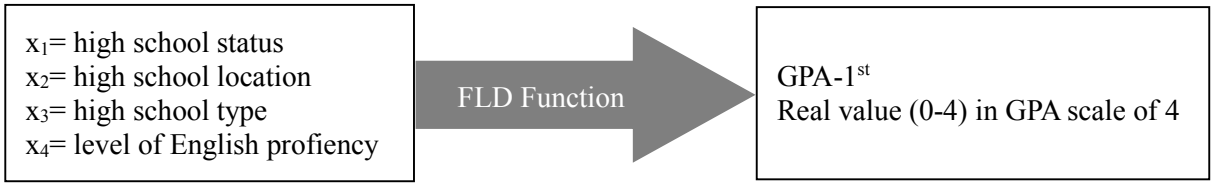

Figure 1: FLD Function Scheme

Contingency table, also known as crosstab, commonly used for displaying interrelations between dependent and independent variables (Bhattacharyya & Johnson, 2010). The example of crosstab is shown in Table 1. The crosstab table can also be used to measure the accuracy of the FLD model by making cross tabulations between the actual (reality) data values and predictive data values.
Example of a crosstab

Table 1: Crosstab Table

		Prediction		Row Count
		Category 1	**Category 2**	
Reality	Category 1	n_{11}	n_{12}	b_1
	Category 2	n_{21}	n_{22}	b_2
Columns count		k_1	k_2	n

The accuracy of the prediction can be measured using Formula 9.
Correctness percentage = ($n_{11} + n_{22}$) / n (9)

METHODOLOGY:

This research was conducted in several stages as follows:

1. Researchers take student data from the academic bureau (which called PUSPINDIKA). As a research limitation, only FTI UKDW student data will be collected for 9 consecutive academic years, from 2008 to 2016. The data from PUSPINDIKA includes high school status, high school type, high school location and GPA-1st.
2. Researchers also take data on English proficiency level for prospective FTI UKDW students from PUSPINDIKA (formerly from PPBA UKDW) for 9 academic years (from 2008 academic year to 2016 academic year).
3. We integrate the two data based on Student Identification Number (called NIM).
4. The researcher grouped the FTI student data based on the admission process for admission of new students. Currently there are two tracks of new student admission at UKDW, namely the academic achievement track and non- academic achievement track (regular track). The difference between these two tracks is the admission process, on the academic achievement track is through academic achievement in high school, while on regular track must take four academic ability tests namely Numerical, Verbal, Spatial and Analogy Ability Tests (Santosa & Chrismanto, 2018).
5. Then after cleaning and integrating student data, the academic achievement data was processed using Fisher Linear Discriminant function. The Fisher Linear Discriminant function model is a *Y* function that can be written as a linear combination of x1, x2, x3, ..., xp, which can be seen in the Formula (7). And look for m as the limit of the FLD function as in Formula (8). Some factors are: a) high school location, which can be categorized as high schools in Java or high schools outside Java, b) type of high school, namely regular or vocational, is also a thought to be studied, c) the status of high school is high school with public status or private status, and d) the level of English proficiency from prospective students is also a factor that influences GPA (Santosa & Chrismanto, 2016). We also use the 4 factors above as independent variables in FLD model and category GPA-1st as independent variable.
6. Nine FLD function models for the case of admission of new students through the academic achievement track are built, in detail can be seen in Table 2.

Table 2: Student Cumulative Data Used to Model FLD

No	Model FLD	Data Training Based on Class
1	Model A	2008
2	Model B	2008 to 2009
3	Model C	2008 to 2010
4	Model D	2008 to 2011
5	Model E	2008 to 2012
6	Model F	2008 to 2013
7	Model G	2008 to 2014
8	Model H	2008 to 2015
9	Model I	2008 to 2016

7. Using nine models of FDL functions to predict the new student's GPA-1st of class of 2017 whose admission is through the academic achievement track.
8. Convert y values to y_{pred} with formulas: $y_{pred} = \begin{cases} 0, \ if \ y < m \\ 1, \ if \ y \geq m \end{cases}$
9. Convert GPA-1st values to y_{real} with formulas:
$$y_{real} = \begin{cases} 0, \ if \ GPA - 1^{st} < GPA - 1^{st} \ average \\ 1, \ if \ GPA - 1^{st} \geq GPA - 1^{st} \ average \end{cases}$$
10. Measuring the accuracy of FLD function prediction results from the nine models against reality observation data by using Crosstab on student admission in 2017.

FINDINGS AND DISCUSSIONS:

Students' personal data we collected are shown in Table 3.

Table 3: Student Cumulative Data That Used To Build Model A-I

No	Class of	Model	Academic Achievement track
1	2008	Model A	63
2	2008 to 2009	Model B	74
3	2008 to 2010	Model C	129
4	2008 to 2011	Model D	273
5	2008 to 2012	Model E	398
6	2008 to 2013	Model F	523
7	2008 to 2014	Model G	613
8	2008 to 2015	Model H	806
9	2008 to 2016	Model I	905
10	2017	Test Data	102

The data we used in this research has some differences compared to our previous research (Santosa & Chrismanto, 2017) because of some data updates. The FLD function models we built are shown in Table 4 and Table 5.

Table 4: Fisher Linear Discriminant Function

Model	FLD Function
A	Y = -3.434 -0.165* (x_1) -0.538* (x_2) + 1.674* (x_3) + 1.195*(x_4)
B	Y = -6.267 + 0.826* (x_1) -0.441* (x_2) +2.890* (x_3) + 1.143*(x_4)
C	Y = -1.582 -0.040* (x_1) - 1.411*(x_2) +1.399* (x_3) + 0.859*(x_4)
D	Y = -0.463 +0.151* (x_1) +1.658* (x_2) +1.122* (x_3) - 0.870*(x_4)

Model	FLD Function
E	$Y = 0.477 - 0.482*(x_1) - 1.486*(x_2) + 0.177*(x_3) + 0.896*(x_4)$
F	$Y = 0.312 - 0.271*(x_1) - 1.369*(x_2) - 0.152*(x_3) + 0.923*(x_4)$
G	$Y = 0.275 - 0.229*(x_1) - 1.222*(x_2) - 0.293*(x_3) + 0.928*(x_4)$
H	$Y = -0.632 - 0.014*(x_1) - 0.269*(x_2) - 0.763*(x_3) + 0.967*(x_4)$
I	$Y = -0.651 - 0.005*(x_1) - 0.429*(x_2) - 0.584*(x_3) + 0.968*(x_4)$

Table 5: Limit Calculation

Model	Limit Calculation	Value
A	$\bar{Y}$ = -3.434 -0.165* (1.57) -0.538 (1.27) + 1.674* (1.02) + 1.195*(2.24)	0.00797
B	$\bar{Y}$ = -6.267 +0.826* (1.59) -0.441 (1.27) + 2.890* (1.01) + 1.143*(2.26)	-0.01165
C	$\bar{Y}$ = -1.582 -0.040* (1.58) -1.411 (1.26) + 1.399* (1.02) + 0.859*(2.33)	0.00539
D	$\bar{Y}$ = -0.463+0.151* (1.62) +1.658* (1.21) + 1.122* (1.04) -0.870*(2.21)	-0.00802
E	$\bar{Y}$ = 0.477-0.482* (1.65) -1.486* (1.20) + 0.177* (1.06) + 0.896*(2.13)	-0.0054
F	$\bar{Y}$ = 0.312-0.271* (1.64) -1.369* (1. 2) - 0.152* (1.07) + 0.923*(2.1)	0.00042
G	$\bar{Y}$ = 0.275-0.229* (1.63) -1.222* (1.22) - 0.293* (1.10) + 0.928*(2.06)	0.00027
H	$\bar{Y}$ = -0.632-0.014*(1.60) -0.269* (1.39)-0.763* (1.16) + 0.967*(1.98)	0.00127
I	$\bar{Y}$ = 0.651-0.005* (1.60) -0.429* (1.39)-0.584* (1.16) + 0.968*(1.98)	-0.01611

Centroid summary results and limits for the Fisher Linear Discriminant function on the various models can be summarized in Table 6.

Table 6: Centroid Summary And The Centroid Function Limit Of FLD With Various Data

Class of	Model	Centroid 1	Boundary (m)	Centroid 2	Correct %
2008	Model A	-0.760	0.00797	0.328	69.84
2008 to 2009	Model B	-0.488	-0.01165	0.194	64.86
2008 to 2010	Model C	-0.627	0.00539	0.302	69.77
2008 to 2011	Model D	0.553	-0.00802	-0.309	67.40
2008 to 2012	Model E	-0.484	-0.0054	0.309	67.09
2008 to 2013	Model F	-0.541	0.00042	0.330	70.17
2008 to 2014	Model G	-0.577	0.00027	0.356	70.47
2008 to 2015	Model H	-0.556	0.00127	0.340	70.35
2008 to 2016	Model I	-0.559	-0.01611	0.349	68.07

The results of the summary of the coefficient of FLD function and the accuracy of the model with actual data can be summarized in Table 7.

Table 7: Table Coefficient Summary And Accuracy Of The FLD Function Model With Various Data Training On Academic Achievement Track

Model	Coefficient					Correct (%)
	Const.	x_1	x_2	x_3	x_4	
2008	-3.434	-0.165	-0.538	1.674	1.195	69.84
2008 to 2009	-6.267	0.826	-0.441	2.890	1.143	64.86
2008 to 2010	-1.582	-0.040	-1.411	1.399	0.859	69.77
2008 to 2011	-0.463	0.151	1.658	0.122	-0.870	67.40
2008 to 2012	0.477	-0.482	-1.486	0.177	0.896	67.09
2008 to 2013	0.312	-0.271	-1.369	-0.152	0.923	70.17
2008 to 2014	0.275	-0.229	-1.222	-0.293	0.928	70.47
2008 to 2015	-0.632	-0.014	-0.269	-0.763	0.967	70.35
2008 to 2016	-0.651	-0.005	-0.429	-0.584	0.968	68.07
Average						**68.67**
Standard Deviation						**1.93**

Table 7 shows that the average of the accuracy percentage is 68.67% with a standard deviation of 1.93. The FLD models were tested to predict student's GPA-1st on student's data from class of 2017 using crosstab and the results are shown in Table 8.

Table 8: Crosstab From Prediction Of The GPA-1st Category Of The 2017 Class Student Use The 9 FLD Functions (Model A To Model I)

Class of	Model	Cell Contents				Correct (%)
		n_{11}	n_{12}	n_{21}	n_{22}	
2008	Model A	28	7	36	31	57.84
2008 to 2009	Model B	15	9	28	50	63.72
2008 to 2010	Model C	19	12	25	46	63.72
2008 to 2011	Model D	0	0	48	54	52.94
2008 to 2012	Model E	21	14	31	36	55.88
2008 to 2013	Model F	20	11	34	37	55.88
2008 to 2014	Model G	20	11	33	38	56.86
2008 to 2015	Model H	20	11	31	40	58.82
2008 to 2016	Model I	20	11	31	40	58.82
Average						58.28
Standard Deviation						3.57

From Table 8, the percentage average of accuracy is 58.28% with a standard deviation of 3.57.

CONCLUSION:

From the research that has been done, then obtained some conclusions as follows:

A. The average accuracy of the 9 classification models for GPA-1st of academic achievement track students was 68.67% with a standard deviation of 1.93.
B. The best accuracy of the FLD function model with training data is found in the model G with a percentage of 70.47%. While the worst accuracy of the FLD function model with training data is found in model B with a percentage of 64.86%
C. While the average accuracy of the prediction results with the 9 models is 58.28% with a standard deviation of 3.57.
D. The accuracy of the best prediction results with the FLD function model is 63.72% using models B and C. While the accuracy of the worst prediction results with the FLD function model is 52.94% using the

model D. This is due to the absence of predictive results of the GPA-1st category match the actual data in the GPA-1st category of students who are below average.

E. In this study, the FLD model performs not well because it has an accuracy average for prediction of only 58.28%.

ACKNOWLEDGEMENTS:

The authors would like to thank Informatics Department of Faculty of Information Technology, DWCU on all funding and infrastructure so that the completion of this research

REFERENCES:

Alexandre, E., Cuadra, L., Gil-Pita, R., & Rosa, M. (2006). Application of Fisher Linear Discriminant Analysis to Speech/Music Classification. *Audio Engineering Society Convention Paper 6678*, (hal. 1-6). Paris, France.

Alhaddad, M. J., Kamel, M. I., Malibary, H. M., Alsaggaf, E. A., Thabit, K., Dahlwi, F., & Hadi, A. A. (June 2012). Diagnosis Autism by Fisher Linear Discriminant Analysis FLDA via EEG. *International Journal of Bio-Science and Bio-Technology Vol.4 No. 2*, 45-53.

Ayala, A. P. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Application, 4*(1), 1432 – 1462.

Berson, A., Smith, S., & Thearling, K. (2011, November 28). *An Overview of Data Mining Techniques*. Diambil kembali dari Data Mining Technique: http://www.thearling.com/text/dmtechniques/dmtechniques.htm

Bhattacharyya, G. K., & Johnson, R. A. (2010). *Statistical Principles and Methods 6th edition.* John Wiley & Sons, Inc.

Dumitra, & Tudor, A. (November 6th - 7th ,2014). Predicting Company Performance By Discriminant Analysis . *Proceedings Of The 8th International Management Conference Management Challlanges for Sustainable Development* ,"(hal. 1173-1180). Bucharest, Romania.

Dunham, M. H. (2002). *Data Mining : Introductory and Advanced Topics.* New York USA: Prentice Hall PTR Upper Saddle River .

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining : Concepts and Techniques.* Morgan Kaufmann Publishers.

Hardle, W. K. (2015). *Applied Multivariate Statistical Analysis 4th edition.* Berlin: Springer-Verlag Berlin Heidelberg.

Jindal, R., & Borah, M. D. (2013). A Survey On Educational Data Mining and Research Trend. *International Journal of Database Management Systems (IJDMS) Vol. 5, No. 3*, 53-72.

Johnson, R. A., & Wichern, D. W. (2008). *Applied Multivariate Statistical Analysis.* New York: Pearson New International Edition.

Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: Introduction to Data Mining Second edition.* John Wiley & Sons, Inc.

Muthii, M. J. (2015). *Predicting Student's Loan Default in Kenya: Fisher's Discriminant Analysis Approach.* Kenya: University of Nairobi.

Ogum, G. (2002). An Application of Discriminant Analysis in University Admission (A Case of the University Medical School, 1975/1976). Dalam G. Ogum, *Introduction to methods of Multivariate Analysis* (hal. 119 – 134).

Okoli, C. (2013). An Application of Discriminant Analysis On University Matriculation Examination Scores For Candidates Admitted Into Anamabra State University. *Journal of Natural Sciences Research, 3*(5).

Rencher, A. C. (2002). *Methods of Multivariate Analysis Second edition.* John Wiley & Sons, Inc. Publication.

Romero, c., & Ventura, S. (2013). Data Mining in Education. *WIREs Data Mining Knowledge Discovery 2013, 3*, 12-27.

Santosa, R. G. (2008). Klasifikasi Indeks Prestasi Mahasiswa TI UKDW dengan Menggunakan Fungsi Discriminant Linier Fisher. *Prosiding snasti 2008 ISBN : 978-979-89683-31-0* (hal. 332-337). Surabaya: STIKOMP Surabaya.

Santosa, R. G., & Chrismanto, A. R. (2016). *Regresi Logistik untuk Prediksi Kategori IP Mahasiswa Fakultas Teknologi Informasi UKDW.* Yogyakarta: Unpublishes.

Santosa, R. G., & Chrismanto, A. R. (2017). Logistic Regression Model for Predicting First Semester Students GPA category Based on High School Academic Achievement. *Researchersworld Journal of Arts, Science & Commerce Volume-VIII Issue-2(1) April 2017*, 58-66.

Santosa, R. G., & Chrismanto, A. R. (2018). Perbandingan Akurasi Model Regresi Logistik untuk Prediksi Kategori IP Mahasiswa Jalur Prestasi dengan Non Jalur Prestasi . *jurnal Teknik dan ilmu Komputer Volume 07 No 25 Januari -Maret 2018*, 107-121.

Santosa, R. G., & Setiadi, H. (2015). *Analisis Faktorial untuk Uji Pengaruh Beberapa Faktor Terhadap Indeks Prestasi Mahasiswa Fakultas Teknologi Informasi UKDW.* Yogyakarta: Unpublished.
