ANALYSIS OF FORECASTING THE NUMBER OF STUDENTS USING VECTOR AUTOREGRESSION COMPUTER BASED

Vivi Andria[1] Bachelor of Computer Science,
Djunaidy Santoso[2], Dipl.Ing.;Magister Computer
Department of Computer Science   Binus University
djunaidyS0533@binus.ac.id

ABSTRACT

Education has become an important factor in a person's life. Therefore, many people who race - the race to get the best education. The course, has become one of the places entrusted by parents in providing additional lessons to their children. Forecasts for the large number of students can be used as a benchmark in determining the success or failure of an educational institution. In addition, can also be used as an assessment of the services rendered, as more and more students are obtained, meaning the greater the trust that is given to perusahaan.Pengumpulan data is done by the secondary, which uses data drawn from the data in the previous period. The data used is the actual data obtained from Valto Education Centre.Data that has been collected will then be processed using Microsoft Excel 2003. The method used was Vector Autoregression[1]. Data processing will generate value for the first forecast future periods.

Keywords*: Vector Autoregression, Time Series, autocorrelation, Forecasting.*

1. INTRODUCTION

Background

The development of education in Indonesia has been growing very rapidly, from kindergarten - age children (kindergarten), elementary school (SD), junior high school (SMP), secondary school (high school), even up to university. One of the most important level of education as basic education is a basic level.

Therefore, much of the race to get the best education for their children. Various methods are used to get a good education and quality such as entering their child's school to the National and International.

To support the education acquired in school, many parents give extra lessons for their children. This is usually done by providing tutoring - additional courses can not be obtained at the school or just to support the lessons they have learned in school in order to become better.

Along with these demands, then indirectly affect the development of business in the field of non-formal education. It is characterized by competition between courses with one another. Proliferation place, where tutoring is also evidence, that education has been a fairly lucrative business.

Problem Formulation :

The quality of education is found to be standard in every company engaged in the field of education. This has become the standard measure of success or failure of a non-formal educational institution in providing the best service for our customers. Many aspects that can be used to determine the success of a company. In this case, the one that is used is with forecasting. Forecasting can be done by counting the number of pupils who gained every month.

The Proposed Solution:

-Limitations

In this Research, the scope is focused on the large number of pupils in Valto Education Centre. The analyzed data is data on the number of students Valto Education Centre every month. This will show a trend that is created by itself. This relationship would be predicted by using Vector Autoregression.
Objectives and benefits In this study, the authors wanted to convey some of the objectives and benefits of the research conducted.

-Research purposes

The general objective of this research is:
1 To determine the existence of the company in the future
2 Predicting the continuation of the company in the future
3 Designing programs help to analyze the large number of pupils in the future

-Benefits

Here are today's benefits for Valto Education Centre include:

1. As input to determine the number of students in the future, so it can be used as a benchmark for the future.
2. As input how good the quality of education provided by the company of students. It can be seen from many or at least the number of students in the company.

3.RESEARCH METHODOLOGY :

In this Research, the authors used two methodologies, that is methodology of analysis and design methodology.
• In the methodology of the analysis of the study of literature and materials taken from the internet that support in solving problems and play a role in the collection of data, that taken from the website. For

the data is used secondary data, where data is extracted from companies Valto Education Centre.
• In the design methodology, will be made the diagrams, the menu, the initial design, modeling, program structure, STD (State Transition Diagram), Unified Modeling Language (UML), and ultimately created by using Visual Studio 2005 as its implementation.

2.REVIEW OF LITERATURE

Regression and Correlation Regression Ref. [1]

Regression can be perceived as a way to estimate a value of the dependent variable with the help of one or more independent variables. The concept of independent variables and the dependent, very important in shaping the proper sense of regression.

The independent variables can be expressed as a value which affects the value of the other, while the dependent variable can be expressed as a value that affected one or more combinations of other values. Although regression can be used to predict the value of a variable dependent on the value of the independent variables, it should be noted that the regression equation is only valid in the data sample interval. In that sense, it is a matter that is not wise to do excessive generalization against a regression equation, for example by including extreme values independent variable is greater than the highest value in the sample

Types of Regression Analysis

Regression analysis grouped ranging from the simplest to the most complex, depending on the destination, which is based knowledge or theory while, not only the origin is determined.

Simple Linear Regression

Simple linear regression aims at studying the linear relationship between two variables. Two variables are divided into the independent variable (X) and the dependent variable (Y). The independent variable is a variable that can be controlled while the dependent variable is the variable that reflects the response of the independent variable.

Multiple regression

Multiple regression is often used to solve regression analysis involving the relationship of two or more independent variables. At first regression developed by econometricians to help predict the result of economic activities in the various segments of the economy. For example, reports about the future of economic forecasting in economic journals (Business Week, Wal Street Journal, etc.), which are based on econometric models with multiple analysis as a tool. One example of the use of multiple regression in agriculture including agricultural scientist uses regression analysis to explore between crops (eg, rice production per hectare) with the type of fertilizer used, the quantity of fertilizers applied, the number of days of rain, temperature, solar radiation, and infection insect.

Curvilinear regression

Curvilinear regression is often used to assess or model the relationship function dependent variable (Y) and the independent variable (X) is not linear. Not linear can be defined when the rate of change of Y as a result of changes in X is not constant for X values specified. Conditions this function is not linear (curvilinear) is often found in many areas. For example in agriculture, can be observed the relationship between rice production level of phosphate fertilizer. In general, rice production will rise rapidly when the administration of phosphate

increased from a low level to moderate. But when phosphate dosing is continued until a high level, then additional doses of phosphate no longer offset the increase in results, on the contrary decreased results. For those cases the relationship is not linear, simple or multiple regression procedure can not be used in the search for patterns of relationships of the variables involved. In this case, the curvilinear regression analysis procedure is the appropriate procedure to use.
Regression With Dummy variables (Doll)

Regression analysis is not only used for quantitative data (eg fertilizers), but can also be used for qualitative data (eg harvest). The qualitative data often indicate the presence of classification (category) specific, often also categorized the independent variable (X) with a nominal measurement classification in the regression equation. For example, if you want to influence regressing pineapple dodol product packaging condition of the sale price. In general, the methods used for the completion is given a value of 1 (one) if the category is there and the value 0 (zero) if a category that does not exist (it could be vice versa, depending on its purpose). In the case of this package, if the packaging is attractive given the value of 1 if it does not pull rated 0. The variable that takes the value of 1 and 0 are called dummy variables and the values given can be used like other quantitative variables. Logistic Regression (Logistic Regression)

When regression with the independent variable (X) in the form of dummy variables, it is categorized as a dummy regression. Logistic regression was used when the dependent variable (Y) in the form of a variable entered the category of classification. For example, the variable Y in the form of two responses that failed (denoted by a value of 0) and successful (denoted by the value of 1). Such conditions are often categorized as a regression with a binary response. As in the multiple

regression analysis, logistic regression for the independent variable (X) can also consist of more than one variable.

Definition Correlation

Throughout the history of mankind, people doing research on the presence and absence of a relationship between two things, phenomena, events or otherwise. Attempts to quantify this relationship is known as measuring the association between two phenomena or events that create the curiosity of researchers.

Correlation is an analytical technique that is included in one measurement techniques associations / relationships (measures of association). Measurement of association is a general term that refers to a group of techniques in bivariate statistics used to measure the strength of the relationship between the two variables. Among the many techniques of measurement of the association, there are two very popular correlation techniques until now, namely the Pearson Product Moment Correlation and Spearman Rank Correlation. In addition to these two techniques, there are also other correlation techniques, such as Kendal, Chi-Square, Phi Coefficient, Goodman-Kruskal, Somer, and Wilson.

Measurement of the association was wearing a numeric value to determine the level of association or the strength of the relationship between variables. Two variables are said to be associated if one variable affects the behavior of other variables. If there is no effect, then the two variables are called independent.

Correlation helpful to measure the strength of the relationship between two variables (sometimes more than two variables) with certain scales, such as the data Pearson must scale interval or ratio; Spearman and Kendal using ordinal scale; Chi Square using nominal data. Strong weak relationship between the measured distance

(range) 0 to 1. Correlations have the possibility of testing hypotheses two-way (two-tailed). Correlation unidirectional if the value of the correlation coefficient was found positive; otherwise if the value of the correlation coefficient is negative, correlation is called not unidirectional. What is meant by the correlation coefficient is a statistical measurement kovariasi or association between two variables. If the correlation coefficient is found not equal to zero (0), then there is dependence between the two variables. If the correlation coefficient was found +1. then the relationship is referred to as a perfect correlation or perfect linear relationship with a slope (slope) is positive.
If the correlation coefficient is found -1. then the relationship is referred to as a perfect correlation or perfect linear relationship with a slope (slope) is negative. In a perfect correlation hypothesis testing is no longer needed, because the two variables have a perfect linear relationship. This means that the variable X affects Y perfectly. If the correlation is equal to zero (0), then there is no relationship between the two variables.

There is some correlation calculation formula daam ststistika, one of which can be written using the formula: This formula is often called Simple Correlation Coefficient.

Another formula that can be used to calculate the value of the correlation coefficient is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

Coefficient of Determination Ref. [1]
The coefficient of determination (R2) essentially measures how far the ability of

the model to explain variations in the dependent variable. The coefficient of determination is between zero and one. R2 value is small means that the ability of the independent variables in explaining the dependent variables is very limited. A value close to the mean of independent variables provide almost all the information needed to predict the variation of the dependent variables Kuncoro (2003, p20). According Suharyadi (2004, p515) states independent variables can explain the dependent variable.

When R2> 0.5 is said to be good or accurate
When R2 = 0.5 is said to be moderate
When R2> 0.5 is said to be lacking

Statistics Durbin – Watson

Statistics Durbin - Watson is one of the statistical tests were used to test existing error autocorrelation of a regression analysis. Autocorrelation test was used to test whether a linear regression model there is a correlation between bullies error in period t with an error in period t-1 (previous). Durbin - Watson after James Durbin and Geoffrey Watson. If et is the error value from the data captured data at period t, then the statistical test that can be used is :

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}.$$

If the value of Durbin Watson statistics are around 2, the error is random. If the Durbin Watson statistics is above 2, meaning there is a positive correlation between the auto error, and vice versa if the value of Durbin Watson statistics are under 2, it means there is a negative correlation between the auto error. Forecasting is good, if the generated random   error.

Analysis of Vector Auto Regression Ref. [1]

Vector Autoregression or VAR typically used to project a coherent system variables of time and to analyze the dynamic impact of interference factors contained in the variable system. Basically, VAR analysis can be paired with a simultaneous equations model, because in the VAR analysis we consider several endogenous variables simultaneously in a model. The difference with the usual simultaneous models are in the VAR analysis of each of the variables other than explained by its value in the past is also influenced by the value of the past all the other endogenous variables in the model were observed. In addition, the VAR analysis is usually no exogenous variables in the model.

The advantages of a VAR analysis include:
1) The method is simple, we do not have to worry to distinguish where the endogenous variables and exogenous variables
2) The estimated simple where the usual OLS method can be used in each equation separately
3) Estimates (forecast) obtained using this method in many cases better than the results obtained using a complex model of simultaneous equations.

In addition VAR analysis also an analysis tool that is very useful both in understanding the reciprocal relationship between economic variables as well as in the establishment of a structured economic model.

Generally VAR with orders p and n endogenous variables at time to t can be modeled as follows:
Where:
Yt = vector of endogenous variables (Y1.y, Y2.t, Yn.t) measuring nx 1
c = Vector sized intercept nx 1
Ai = xn matrix size parameter
t = vector of residual (1t, 2t, ... nt) measuring nx 1

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

Basically, VAR analysis include: The unit root test (Unit Root Test) The unit root test is used to see if the observed data is stationary or not. This test is actually just a complement of VAR analysis, given the purpose of VAR analysis is to assess the reciprocal relationship between variables observed, and not a test for the data. However, if the observed data is stationary, it will improve the accuracy of the analysis of VAR.

Hypothesis (Hyphothesis Testing), which consists of:

Likelihood Ratio Test

Likelihood Ratio Test is used to test hypotheses about what is the amount of lag corresponding to model the observed. Granger Causality Test

These tests examine whether an independent variable (independent variable) improve the forecasting performance of the dependent variable (dependent variable).

Accounting Innovation

Basically, this test was used to test the dynamic structure of the system variables in the model were observed, which is reflected by the variable innovation (innovation variable). In other words, this test is a test of the variables of innovation (innovation variable). This test consists of:

The Impulse Responses:

To see the effect of the turmoil (shock) a standard deviation of variable invovasi the present value (current time values) and value (future values) of endogenous variables that were able to model the observed.

The Cholesky Decomposition:

The Cholesky Decomposition or commonly referred to as The Variance Decomposition

provide information about the variable that is relatively more important innovation in the VAR. Basically, this test is another method to describe dynamic systems contained in the VAR. This test is used to compile estimates of error variance of a variable, that is how big the difference between the variance before and after shock, shock that comes either from ourselves and shock of other variables.

Data Used

The data used is data on the number of students Valto Education Centre commencing from March 2005 to October 2008.

Mindset

As a company established in the field of care and services, the customer trust becomes so determines the existence of the company. Likewise with the company Valto Education Centre, as a company engaged in the field of education services, the number of students who are there every month become a benchmark for the success or failure of the company. Forecasts for the large number of students Valto Education Centre is necessary to know how to live in the future. For this forecast, used historical data on the number of students studying Valto Education Centre in the month - previous month.

The method used is the method of Vector Autoregression. This method can be used to predict subsequent data. Data months - the previous month will be collected will then be used as a benchmark for calculating the value forecast for further data. Predictive value is best to have an error (error) is random. It can be seen from the Durbin-Watson statistics produced, which is located around a value of 2.

Research Hypothesis
The hypothesis on existing research in this study are:

Are the existing data in the past can be used to predict the data in the future.

Data Analysis Techniques   Ref. [1]

The collected data will be estimated using Vector Autoregression. Previous data is tested first by using the Time Lag. This is to determine if the previous data have any impact on data going forward. Time Lag After getting the best value to be used, it will get the regression equation that supports predictive value.

Number of students in the future greatly affected by the number of students at one month earlier and was followed by the data in the previous two months
.
From data on the number of students before, obtained the best autoregresi equation is: Model autoregresi obtained can already be used to predict the number of students for one period to come.

The correlation coefficient to be used, can be formulated as follows.

$$\bar{Y}_t = c + A_1 Y_{t-1} + A_2 Y_{t-2}$$

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

4. RESULTS   AND   DISCUSSION   :

Presentation   of   Data   Research
Following is data on the number of students Valto Education Centre for the period from March 2005 to October 2008. The following data is actual data to predict the data in the coming period. The total number of actual data is 44.

Table 4.1 Data student period March 2005 to October 2008.

Period 1 2 3 4 5 6 7 8   9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

Yt   4 6 8   7 9 13 12 11 12 9   12 13 15 17 16 12 14 16 15 17 15   14 13 14 15 14 16 11 11 14 15 17 18 15 20 23

Period 37 38 39 40 41 42 43 44

Yt   28 27   29 23 24 30 36 38
Figure 4.1 below, will display a line chart for all of the data obtained, it can be seen the pattern (Trend) is formed.

16

Figure 4.1 The line chart to preliminary data the number of students. From the picture above, it can be seen that the data is likely to increase for a certain period of time. Existing data can be quite stable to see the trend is formed.

Table 4.2 Autocorrelation Table

Time Lag	1	2	3	4	5
Autocorreation	0.98419	0.75548	0.75548	0.574	0.423

From Table 4.2, we can see the result of autocorrelation is formed. The autocorrelation value is to use Time Lag = 1. This indicates that the data for one period that will come will be influenced mainly by the data one month earlier, followed by data on the two previous months.

Data Processing

The number of students Valto Education Centre have been obtained will be analyzed using the methods Vector Autoregression. The analysis was conducted to obtain predictive value of the number of students. There are several steps that must be done to get a good predictive value, such as by finding the equation autoregresi formed, counting the number of errors, analyzing the equation formed.

Determining Time Lag

To obtain the best regression equation, it must be determined beforehand Time Lag to be used. Autoregresi equations used are:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

Time Lag = 1

Below can be seen the data using Time Lag = 1. Data taken from previous data.

Table 4.3 Data using the Time Lag = 1

Period 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

Yt 4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15 20 23

Yt-1 4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15 20

Period 37 38 39 40 41 42 43 44

Yt 28 27 29 23 24 30 36 38

Yt-1 23 28 27 29 23 24 30 36

From the data in Table 4.3, it will get the value of the coefficient of determination 0.86429. This means that the diversity of the data can be explained by the model of 86.429% and the equation autoregresi

$$\bar{Y}_t = 0.515439 + 1.016062 Y_{t-1} - 0.00079 Y_{t-2}$$

formed as follows:

$$\bar{Y}_t = 0.6877 + 1.0065Y_{t-1}$$

Time Lag = 2

Below can be seen the data using Time Lag = 2.

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t,$$

Table 4.4 The number of students using the Time Lag 2

Period 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

Yt   4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15 20 23

Yt-1   4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15 20

Yt-2   4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15

Period 37 38 39 40 41 42 43 44

Yt   28 27 29 23 24 30 36 38

Yt-1 23 28 27 29 23 24 30 36

Yt-2 20 23 28 27 29 23 24 30

From the data in table 4.4, meaning the value of the coefficient of determination obtained 0.857735. This means that the diversity of the data can be explained by the model of 85.77% and autoregresi equation as follows:

Time Lag = 3

Below   can   be   seen   the   data   using   Time   Lag   =   3.

Tabel4.5 The number of students using the Time Lag = 3

Period 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

Yt   4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15 20 23

Yt-1   4 6 8 7 9 13 12 11 12 9 12   13 15 17 16 12 14 16 15 17   15 14 13 14 15 14 16 11 11 14 15 17 18 15 20

Yt-2   4 6 8 7   9 13 12   11 12 9   12 13 15 17 16 12 14 16 15   17 15 14 13 14 15 14 16 11 11 14 15 17 18 15

Yt-3   4 6 8 7   9 13 12   11 12 9   12 13 15 17 16 12 14 16 15   17 15 14 13 14 15 14 16 11 11 14 15 17 18

Period 37 38 39 40 41 42 43 44

Yt   28 27 29 23 24 30 36 38

Yt-1   23 28 27 29 23 24 30 36

Yt-2   20 23 28 27 29 23 24 30

Yt-3   15 20 23 28 27 29 23 24

Data From table 4.5, meaning obtained determination coefficient value 0.85495. It Singer Means Data Model can be explained by diversity of 85.495% and autoregresi equation as follows:

$$\ddot{Y}_t = 0.12233 + 1.00915 Y_{t-1} + 0.237055 Y_{t-2} + 0.17905 Y_{t-3}$$

Time Lag = 4 The singer can be Seen Data of Students using Timelapse = 4.

Table 4.6 Period of data using TimeLapse Student BY = 4

Period 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

Yt   4 6 8 7 9 13 12 11 12 9 12 13 15 17 16 12 14 16 15 17 15 14 13 14 15 14 16 11 11 14 15 17 18 15 20 23

Yt-1   4 6 8 7 9 13 12 11 12 9 12   13 15 17 16 12 14 16 15 17   15 14 13 14 15 14 16 11 11 14 15 17 18 15 20

Yt-2   4 6 8 7   9 13 12   11 12 9   12 13 15   17 16 12 14 16
15   17 15   14   13   14   15   14   16   11   11   14   15   17   18   15

Yt-3   4   6 8 7   9 13 12   11 12 9   12   13 15   17 16 12 14
16 15   17 15   14   13   14   15   14   16   11   11   14   15   17   18

Yt-4   4   6 8 7   9 13 12   11 12 9   12   13 15   17 16 12 14
16 15   17 15   14   13   14   15   14   16   11   11   14   15   17

Period 37 38 39 40 41 42 43 44

Yt   28 27 29 23 24 30 36 38

Yt-1   23 28 27 29 23 24 30 36

Yt-2   20 23 28 27 29 23 24 30

Yt-3   15 20 23 28 27 29 23 24

Yt-4   18 15 20 23 28 27 29 23

From the data in Table 4.6, the values obtained determination coefficient of 0.858076. This means that the diversity of the data can be explained by the model of 85 807% and autoregresi equation as follows:

$$\overline{Y}_t = 0.126654 + 0.981439Y_{t-1} - 0.0832Y_{t-2} - 0.013466Y_{t-3} + 0.293407Y_{t-4}$$

Time Lag = 5

Below can be seen the student data using Time Lag = 5.

Table 4.7 The number of students using the Time Lag = 5

Period 1 2 3 4 5 6 7 8   9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

Yt   4 6 8   7 9 13 12 11 12 9   12 13   15 17 16 12 14 16 15 17 15   14 13 14 15 14
16 11 11 14 15 17 18 15 20 23

Yt-1   4 6   8 7 9   13 12 11 12 9 12   13 15 17   16 12 14   16 15 17   15 14
13 14   15 14   16 11   11 14 15   17 18   15   20

Yt-2   4   6 8 7   9 13 12   11 12 9   12   13 15   17 16 12 14 16
15   17 15   14   13   14   15   14   16   11   11   14   15   17   18   15

Yt-3   4   6 8 7   9 13 12   11 12 9   12   13 15   17 16 12 14
16 15   17 15   14   13   14   15   14   16   11   11   14   15   17   18

Yt-4   4 6 8 7 9 13 12 11 12 9   12 13 15 17 16 12 14
16 15   17 15   14 13   14   15   14   16   11   11   14 15   17

Yt-5   4 6 8 7 9 13 12 11 12 9   12 13 15 17 16 12
14 16 15   17 15   14 13   14   15   14   16   11   11   14 15

Period 37 38 39 40 41 42 43 44

Yt   28 27 29 23 24 30 36 38

Yt-1 23 28 27 29 23 24 30 36

Yt-2 20 23 28 27 29 23 24 30

Yt-3 15 20 23 28 27 29 23 24

Yt-4 18 15 20 23 28 27 29 23

Yt-5 17 18 15 20 23 28 27 29

From the data in Table 4.7, then obtained determination coefficient value 0.86914. This means that the diversity of the data can be explained by the model of 86.91% and autoregresi equation as follows:

$$\hat{Y}_t = -0.401417 + 0.911132Y_{t-1} - 0.071563Y_{t-2} - 0.13024Y_{t-3} + 0.05267Y_{t-4} + 0.3447281Y_{t-5}$$

Determining Equations Autoregresi

In Table 4.3 to 4.7, obtained the regression equation is best regression equation by using the Time Lag = 2. Because the Time Lag = 2, has a large coefficient determination, ie 0.8577 and also autocorrelation great value, namely 0.75548. Thus, in determining the value of forecasting for the data to come, can use autoregresi equation as follows:

$$\hat{Y}_t = 0.515439 + 1.016062Y_{t-1} - 0.00079Y_{t-2}$$

Counting Error Each Data

After the regression equation to be used, then the next step is to look for the error value of any data that is around equation autoregresi earlier. The formula for calculating the square of each error and the error of data are: Here is a diagram drawing point for error (et) each data.

$$e_t = Y_t - \bar{\bar{Y}}_t$$

$$e_t^2 = (Y_t - \bar{Y}_t)^2$$

From Figure 4.2 below, it can be seen that the observation of one and the other are not mutually affect or be free.
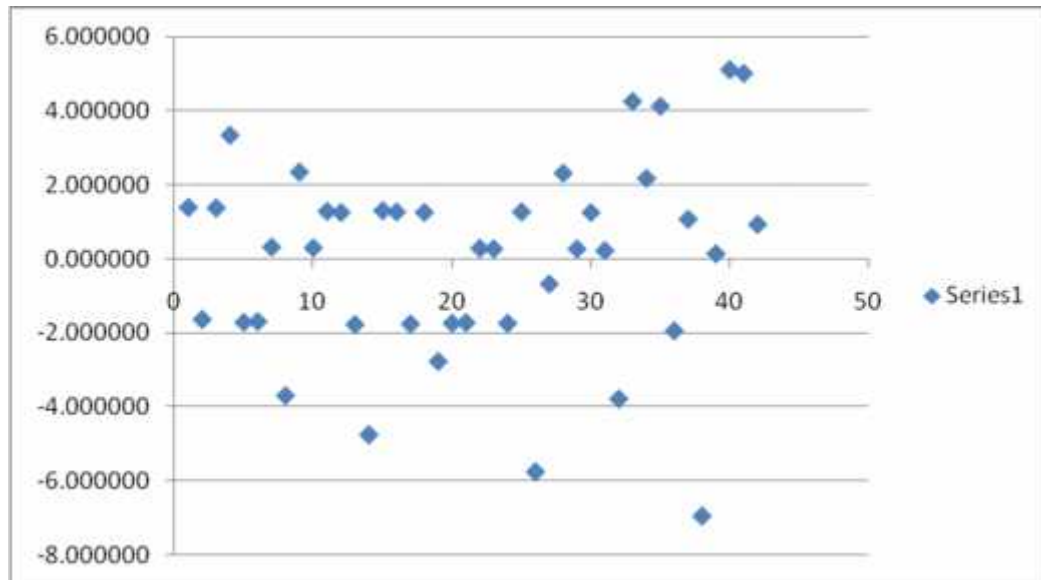

Diagram Figure 4.2 grade point error (et) each data.

$$D - W = \frac{\sum_{t=1}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n}(e_t)^2}$$

Having obtained the error of each data then will be calculated value of Durbin Watson statistics. Durbin Watson statistic can be obtained by using the formula: If the value of Durbin Watson statistics are around 2, the error is random. If the Durbin Watson statistics is above 2, meaning there is a positive correlation between the auto error, and vice versa if the value of Durbin Watson statistics are under 2, it means there is a negative correlation between the auto error. Forecasting is good, if the generated random error.

$$\hat{Y}_t = 0.515439 + 1.016062 Y_{t-1} - 0.00079 Y_{t-1}$$

Table 4.8 Calculation of the value of statistics Durbin - Watson's models

Period	Yt	estimators y	et	et^2	(et-et-1)^2	d
1	4					
2	6					
3	8	6.608666	1.3913336	1.935809197	1.935809197	0.006469871
4	7	8.639218	-1.639218	2.687035154	9.184242104	0.030695617
5	9	7.621583	1.3784173	1.900034226	9.10612183	0.030434523
6	13	9.654494	3.3455064	11.19241324	3.869439666	0.012932459
7	12	13.71717	-1.717169	2.948670818	25.63068672	0.085663001
8	11	12.69796	-1.697961	2.883072712	0.00036895	1.23311E-06
9	12	11.68269	0.3173144	0.10068842	4.061336252	0.013573817
10	9	12.69953	-3.699534	13.6865539	16.13507322	0.053926717
11	12	9.650561	2.3494388	5.519862581	36.5900751	0.12229152

12	13	12.70111	0.2988928	0.089336892	4.204738912	0.014053098
13	15	13.71481	1.28519	1.651713316	0.972782198	0.003251237
14	17	15.74615	1.2538521	1.572145012	0.000982065	3.28226E-06
15	16	17.7767	-1.776699	3.156660697	9.184242104	0.030695617
16	12	16.75906	-4.759064	22.64869248	8.894500166	0.02972724
17	14	12.6956	1.3043981	1.701454332	36.76557527	0.122878078
18	16	14.73087	1.2691278	1.610685361	0.001243992	4.15768E-06
19	15	16.76142	-1.761424	3.102613298	9.184242104	0.030695617
20	17	15.74379	1.2562115	1.578067288	9.10612183	0.030434523
21	15	17.7767	-2.776699	7.710059462	16.26437004	0.054358854
22	14	15.743	-1.743002	3.038056136	1.068530182	0.003571247
23	13	14.72851	-1.728513	2.987756472	0.000209939	7.01657E-07
24	14	13.71324	0.2867629	0.08223298	4.061336252	0.013573817
25	15	14.73009	0.2699143	0.072853711	0.000283878	9.48777E-07
26	14	15.74536	-1.745361	3.046286625	4.061336252	0.013573817
27	16	14.72851	1.2714872	1.61667972	9.101375885	0.030418661
28	11	16.76142	-5.761424	33.19400255	49.46183523	0.165311577
29	11	11.67954	-0.67954	0.461774245	25.82554424	0.086314255
30	14	11.68347	2.3165279	5.366301582	8.976421337	0.030001037
31	15	14.73166	0.2683413	0.072007066	4.195068311	0.014020777
32	17	15.74536	1.2546385	1.574117867	0.972782198	0.003251237
33	18	17.7767	0.2233006	0.049863166	1.063657911	0.003554962
34	15	18.79119	-3.791189	14.37311129	16.11612398	0.053863385
35	20	15.74222	4.2577844	18.1287282	64.78596735	0.216527964
36	23	20.82489	2.175114	4.731121028	4.337515984	0.014496866
37	28	23.86914	4.1308598	17.06400261	3.824941492	0.012783737
38	27	28.94709	-1.947092	3.791166407	36.94149531	0.123466039
39	29	27.9271	1.0729028	1.151120353	9.12036709	0.030482134
40	23	29.96001	-6.960008	48.44171269	64.52765696	0.215664637
41	24	23.86206	0.137938	0.0190269	50.38083917	0.168383077
42	30	24.88284	5.117157	26.18529582	24.79262163	0.082862016
43	36	30.97843	5.0215703	25.21616823	0.009136819	3.05371E-05
44	38	37.07008	0.9299159	0.864743653	16.74163537	0.055953972
Total	**718**	708	-4.8E-14	299.2036977	601.4586345	2.010197866

From table 4.8 above, it can be seen that the value of Statistics Durbin - Watson produced is 2.0101. Values obtained statistics Durbin Watson, still be around a value of 2, so that the resulting error is random or it means there is no correlation between an individual error.

$$\bar{Y}_t = 0.515439 + 1.016062Y_{t-1} - 0.00079Y_{t-2}$$

Data Analysis

In calculating the number of students, the data used is the data of the past. From these calculations, equation obtained following autoregresi

Thus, forecasting the value obtained for student data at 1 month later, 45 months is 39.1 students.

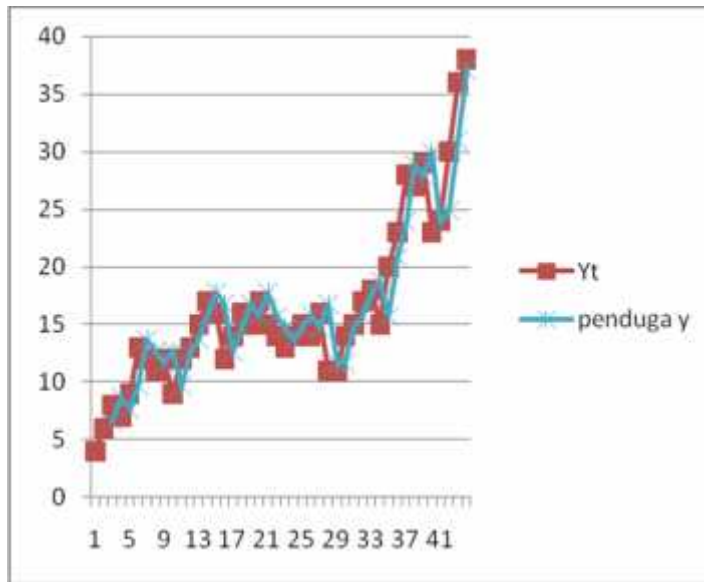$$Y_{45} = 0.515439 + 1.016062(38) - 0.00079(36) = 39.1$$

Figure 4.3 The graph comparison of actual data with the data estimator.
Note : penduga=estimator/predictive/forecast

Discussion :

From the results above, it can be seen predictive value for the future. By using Vector Autoregression, and with the help of software (software) of the C # programming language of Visual Studio 2005 package, the program can calculate forecast values for the data to come up with using data that already exists. With the regression equation generated, would make the resulting error to a minimum, so that the resulting predictive value was also good to get a forecast. Software (software) programming language Visual C # of Visual Studio 2005 package, have helped to speed up the analysis process of forecasting the number of students, through the analysis of Vector Autoregression.
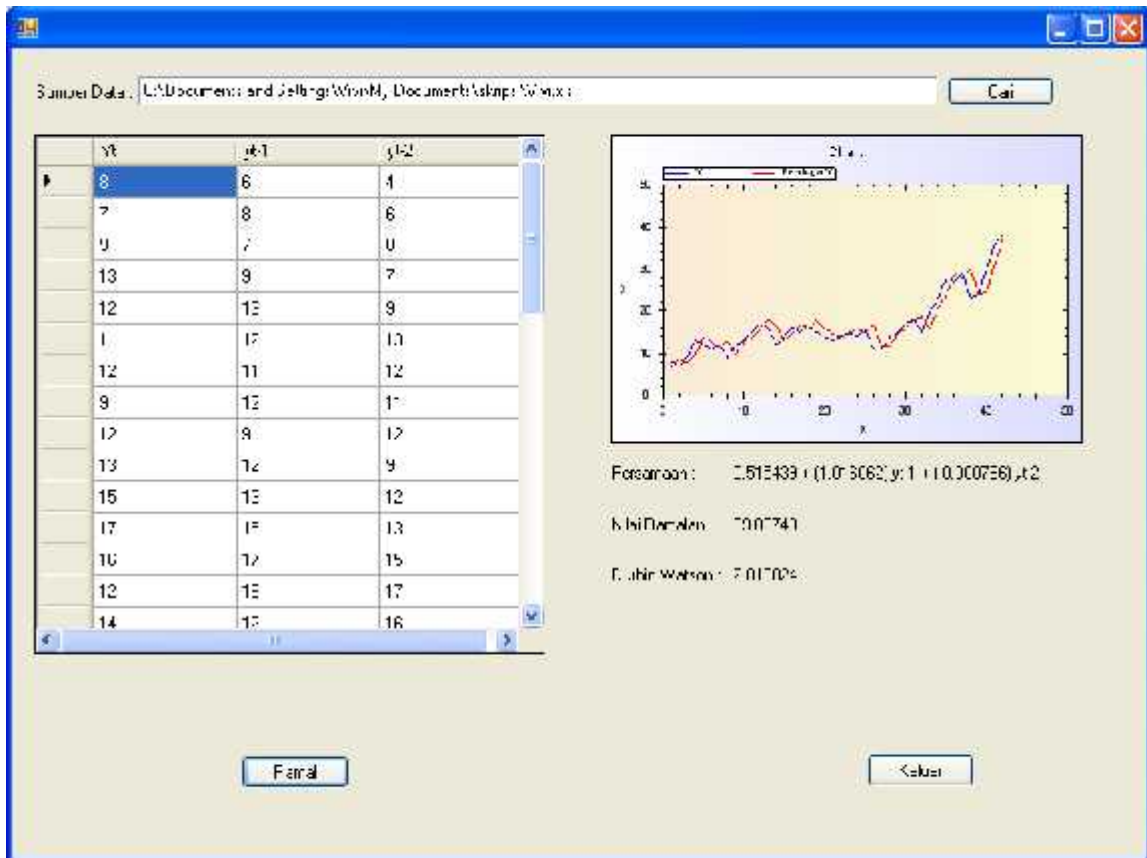
Figure 4.4 Display Screen Design Program

See the Meaning of the Figure above :

Sumber data= data resource;   cari = search;   persamaan = equation; nilai Ramalan= Forecast value;
Dubin Watson = Person who make Analysis statistic AutoRegression;   Ramal = Forcasting;   keluar= exit

5.CONCLUSION :

From the research forecasting the number of these students, summed up some of the following:

$$Y_i = 0.515439 + 1.016062 Y_{t-1} - 0.00079 Y_{t-2}$$

a. Forecasting calculations performed may use the data in the past. From these calculations, equation obtained autoregresi best to forecast data 1 coming period are:

b. Programs designed for this Autoregression Vector forecasting method will allow a user to obtain the value of
forecasting the future
c. Value of forecasting the number of students is influenced by the number of students, especially that of the one previous period and was followed by two previous periods.

Suggestions

Here are suggestions - suggestions are given for further research on forecasting the number of students:
a. If you want to do further analysis, it can be added factors - other supporting factors such as, the number of the company's revenue, the number of students in the schools about the location of the course, as well as the factors - other factors relating to the number of students.

b. The analyzed data can be extended timescales for data retrieval, to get an idea of the wider population.

6.REFERENCES

[1] Assauri,Sofyan. Technique and Its Application in Economic Forecasting Method and Usaha.Fakultas World Economics,1984. University of Indonesia, Yogyakarta.

[2] Booch, Grady, James Rumbaugh, dan Ivar Jacobson. *The Unified Modeling Language User Guide*. 1999. Addison-Wesley. Longman Ltd. Germany.

[3] Deitel, Harvey M, dan Paul J. Deitel. *C# How to Program. 2003*. Prentice Hall, Inc., New Jersey.

[4] Farthansyah.Advanced Database Systems Books Database.2004. Informatics, Bandung

[5] Kuncoro,M. *Research Methods for Business and Economics.2003*. Erlangga

[6] Litterman.R.B. *Techniques of Forecasting Using Vector Autoregressions*.1979. Working Paper No. 115.

[7] Macho.F.J.F. *Forecasting and Interpolation Using Vector Autoregressions with Common Trends.1987.* Journal of Statitics No. 6/7.

[8] Pappalarado,C. and G.Piras. *Vector-Auto-Regression Approach to Forecast Italian Imports.2004.* Working Paper No. 42.

[9] Pressmann,R.S. *S*oftware engineering. Translated by LN Harnaningrum.Andi. 2002. Yogyakarta.

[10] Sentosa, Isap. Human-Computer Interaction: Theory and Practice.1997. First Edition. Andi, Yogyakarta.

[11] Shneiderman, Ben. *Designing The User Interface :* Strategies For Human Computer Interaction. 1998. Third Edition. Addison – Wesley.

[12] Sudjana. *Regression Analysis and Correlation Technique For Researchers*.2003. Issue 3. Tarsito. Bandung.

[13] Sugiarto,H. Forecasting Bisnis.2000. Gramedia Pustaka Utama, Jakarta.