Support Vector Regression for GPA Prediction

K E Dewi1*, N I Widiastuti2
1,2Informatic Engineering Department, Universitas Komputer Indonesia, Indonesia
Email: *kania.evita.dewi@email.unikom.ac.id

Abstract. This study aims to predict student GPA. This research began by collecting data. The features used in predicting GPA are semester 1 and semester 1 IP grades. The process of GPA prediction uses SVM regression, Linear Regression, and Simple Linear Regression. Based on testing with normalized data, the smallest error is obtained by the SVM regression method with Kernel RBF which is equal to 0.1505. Whereas by using standardized data, the smallest error is obtained by using the SVM regression improve method with the Kernel RBF, which is 0.1487. Based on this research, in order to obtain prediction results that are closer to the actual values, it is better to standardize the data first and to predict the process using the SV Regression Improve method using the Kernel RBF.

1. Introduction
The level of education in Indonesia consists of early childhood education, basic education, secondary education, and higher education, according to Law Number 20 of 2003[1]. Most secondary education graduates hope to continue to tertiary education. The higher education level is different from the previous levels. Because at this level has begun to be more direct according to one's abilities. It is not uncommon for someone to choose the wrong major. This can be seen from the number of students choosing to change majors or even choosing not to continue studying[2].

GPA prediction at the beginning of the academic year can help students predict learning success. By predicting a student's GPA, it can help monitor student progress [2,3]. Predicting the CPI has been done by many researchers beforehand. In 1975, Brad S Chissom [4] predicted a student's GPA using a multiple regression method based on SAT scores and grades in high school. In the research of Chris Verhoven and Arie Van Staveren [5], a system for grade prediction was made. Various methods have also been used, some of which use factorization machines[3], Restricted Boltzman Machines [2,6] Decision trees [7], linear regression [8,9] Support Vector Regression [10]. In Huang [10] research predicting the value of dynamics courses, using statistical subjects, calculus 1, calculus 2, physics, and 3 quiz values, obtained using SVM better than linear regression, multilayer perceptron neural network.

In this study, the GPA will be predicted using the Support Vector Machine Regression (SVR) method. By using 1462 data on the value of students who have graduated. Before being predicted using the SVR method the data is normalized and standardized first. Based on the test results obtained that the standardized data has a smaller error when compared to the normalized data first. The kernel that produces the smallest error is the Radial Base Function (RBF). The conclusion from this study the prediction of GPA can be done using SVR.
2. Method
The stages of this research are:

2.1. Study the literature and formulate problems
Researchers study journals about GPA predictions and researchers also study machine learning methods for prediction. This stage ends with the resulting problem formulation for this research.

2.2. Collecting data
At this stage, the researcher coordinates with PTSI to obtain data. The data used in this study are eight-semester 1 course scores, semester one GPA and cumulative GPA of students who have graduated.

2.3. Processing data
In this study to predict cumulative GPA, the SVR method is used, according to Neeraj [11], it is said that the data should be normalized first. Normalization here makes the data range between 0 and 1. Also, the data is standardized to match the Gaussian distribution.

2.4. SVR training and testing
The method used to predict the GPA in this study includes the supervised method, so the data is divided into 2 namely training data that will be used to make prediction models and test data to test the models generated from the training process.

2.5. Conclusion
In this study to see the performance of the machine using an error. The error values used in this study are Mean absolute error (MAE) and root mean squared error (RMSE).

3. Results and Discussion
Research on the prediction of GPA has been carried out. Most research that has been done only classifies GPA [12][8] or even just predicts the value of a course [7]. While what is done in this study is to predict the GPA of a student based on eight semesters 1 grades and semester 1 GPA. In this study, it is assumed that eight-course scores in the first semester and the first semester GPA can demonstrate the adaptability of students with higher education. This is because in semester 1 there are courses that are the basis of the Information Technology major, namely Calculus 1 and Algorithms and Programming.

In this study to predict the Cumulative GPA using the Support Vector Regression (SVR) method. The use of SVR in predicting has been done a lot. As in Yang research [13] that uses SVM in predicting the price of gold, which strongly advocates the use of SVM. There are also studies of Li [14] which are protein folds prediction and many more. Before applying the SVR method in predicting, according to Neeraj Kumar [11] the data must be normalized first. So that the data is valued between 0 and 1. To normalize in this study using equation 1 where \(s_k\) is the corresponding value and \(s\) is the value to be normalized[15].

\[
s' = \frac{s - \min(s_k)}{\max(s_k) - \min(s_k)}
\] (1)

In addition to normalizing data first, before using SVR we also need to know the characteristics of the data encountered. Each data has different characteristics. There is data that can be separated linearly and there is data that cannot be linearly separated. Because in this study using nine attributes, it is difficult to know that the data can be linearly separated or not. So in this study, it is assumed that the data cannot be separated linearly. Therefore, in predicting the cumulative performance index the kernel function is used to transform data. In this research, the kernel functions that are used [Grab your reader’s attention with a great quote from the document or use this space to emphasize a key point. To place this text box anywhere on the page, just drag it.]
Sed is the polynomial kernel and the Radial Basis Function (RBF) kernel[16]. In the polynomial kernel, \( x_i \) is the i-th data and \( x_j \) is the j-th data, while \( d \) is the degree of a polynomial, the polynomial kernel equation can be seen in equation 2. The equation for the RBF kernel also has the same variable with the addition of \( \gamma = \frac{1}{\sigma} \) as a value of data variation, the RBF kernel equation can be seen in equation 3

\[
K(x_i, x_j) = (x_i^T x_j + 1)^d
\]

\[
K(x_i, x_j) = \exp \left( -\frac{1}{2} \gamma \|x_i - x_j\|^2 \right)
\]

Tests carried out using Weka 3.8.4. To find out the performance of the algorithm, what is considered is the error results obtained from each method. The errors used are Mean Square Error (MSE) and Root Mean Squared Error (RMSE). The first test scenario, using normalized data, in making predictive models using SVR kernel polynomials with values \( d = 1, 2, \) and 3. The results of testing using the polynomial kernel can be seen in table 1. Based on table 1 can be seen as the smallest error obtained at the time \( d = 1 \) (linear kernel) with an error value of 0.1512 for the MAE value and 0.1896 for the RMSE value. The second test scenario, using normalized data, predictive modeling using SVR with RBF kernel, with \( \gamma \) value = 0.1; 0.01; 0.001; and 0.0001. The results of the second test scenario can be seen in table 2. Based on table 2, it was found that \( \gamma = 0.1 \) gets the best error, with an error of 0.149 for the MAE value and 0.1862 for the RMSE value. So to predict the cumulative achievement index, based on tables 1 and 2 it is better to use the RBF kernel.

**Table 1. Test Results using Normalized Data And Polynomial Kernels (\( d = 1, 2, \) and 3)**

| d   | MAE   | RMSE  |
|-----|-------|-------|
| 1   | 0.1512| 0.1896|
| 2   | 0.1521| 0.1892|
| 3   | 0.1566| 0.196 |

**Table 2. Test Results using Normalized Data And RBF Kernels (\( \gamma = 0.1; 0.01; 0.001; 0.0001 \))**

| \( \gamma \) | MAE   | RMSE  |
|--------------|-------|-------|
| 0.1          | 0.149 | 0.1862|
| 0.01         | 0.1506| 0.1885|
| 0.001        | 0.1603| 0.1987|
| 0.0001       | 0.2119| 0.2591|

This research also shows the effect of pre-data processing by standardizing the data. The process of standardizing data is done by changing values into standard numbers. The standardization process is to change the data set to an average of 0 and variance 1. If \( x \) is the value to be changed, and \( \bar{x} \) is the average of the related data (in 1 column/attribute) and \( \sigma \) is the standard deviation of the related data (in 1 column/attribute). Then the standardization process can be seen in equation 4[17].

\[
z = \frac{x - \bar{x}}{\sigma}
\]

Using standardized data, three and four test scenarios are performed. The third test scenario is to use standardized data, prediction models are made using SVR with a polynomial kernel, with values \( d = 1 \) and 2. Scenario 3 test results can be seen in table 3. Based on the test results that can be seen using the polynomial kernel obtained by the value the smallest error at \( d = 1 \) is 0.1512 for MAE values and 0.1896 for RMSE values. While testing scenario 4, the prediction model is created using SVR with the RBF kernel.
kernel. The results of testing with scenario 4 can be seen in table 4. Based on table 4 can be seen with the RBF kernel the smallest error value is no longer at gamma = 0.1 but changes to gamma = 0.01. The error value obtained at 0.01 gamma was 0.1487 for the MAE value and 0.1861 for the RMSE value. This value is the smallest error value obtained from all test scenarios that have been carried out. Then in this study, the 5th test scenario will be continued, by comparing the results of the RBF kernel SVR with 0.01 gamma with the predicted results using linear regression and simple linear regression methods. Because these two algorithms are often and the simplest in predicting a value. The results of testing scenario 5 can be seen in table 5.

Table 3. Test Results using Standardized Data And Polynomial Kernels (d = 1 and 2)

| d  | MAE   | RMSE  |
|----|-------|-------|
| 1  | 0.1512| 0.1896|
| 2  | 0.2222| 0.2772|

Table 4. Test Results Using Standardized Data And RBF Kernels (γ =0.1; 0.01; 0.001; 0.0001)

| γ  | MAE   | RMSE  |
|----|-------|-------|
| 0.1| 0.153 | 0.191 |
| 0.01| 0.1487| 0.1861|
| 0.001| 0.1514 | 0.189 |
| 0.0001| 0.1779 | 0.2194|

Another name for the RBF kernel is the Gaussian function [16]. In this study, the data is also standardized. The standardization process in this study is intended to make the data have an average of 0 and the variance of the data is 1. So that the standardized data follows the Gaussian distribution. The results of testing using standardized data can be seen in table 5 obtained better results for the RBF kernel.

Table 5. Test Results using Standardized Data

|                      | MAE   | RMSE  |
|----------------------|-------|-------|
| SVR RBF              | 0.1487| 0.1861|
| Linear Regression    | 0.1511| 0.1894|
| Simple Linear Regression | 0.1632| 0.2026|

Based on the test results above it can be seen that the results of this study are in line with research Huang [10], which uses smaller SVR errors than Linear Regression. In contrast to the results of Muflikah's research [18], in his research SVM is used to calciﬁy sentiment analysis, the polynomial kernel is better than the RBF kernel. This is because in this study the data has gone through a standardization process. So the data has been made following the same Gaussian distribution as the RBF kernel function. In this study do not compare with the Decision trees as in the study of Mashael et al [7] because in this study machine learning is used to predict not to classify.

4. Conclusion

Based on the result and discussion, it is found that by standardizing the data by changing the average data to 0 and the variance to 1, the prediction error is smaller. This means that standardized input data can further assist the machine in creating models that are compatible with the data. Standardized data better meets the RBF kernel assumptions, so using the RBF kernel the smallest error is obtained.

Acknowledgement

We thank System Information and Technology Development (Pengembangan Teknologi & Sistem Informasi) Universitas Komputer Indonesia for providing the data in this study.
References

[1] Keswoyo B, 2003. Tahun 2003 Tentang Sistem Pendidikan Nasional UU_no_20_th_2003. pdf

[2] Iqbal Z, Qadir J, Mian A N and Kamiran F 2017 Machine learning based student grade prediction: A case study arXiv Prepr. arXiv1708.08744

[3] Sweeney M, Lester J and Rangwala H, 2015. Next-term student grade prediction 2015 IEEE International Conference on Big Data (Big Data) pp 970–5

[4] Chissom B S and Lanier D, 1975. Prediction of first quarter freshman GPA using SAT scores and high school grades Educ. Psychol. Meas. 35 461–3

[5] Verhoeven C and Van Staveren A ,2004. An automated grade prediction system Eur. J. Eng. Educ. 29, PP. 251–8

[6] Iqbal Z, Qayyum A, Latif S and Qadir J, 2019. Early Student Grade Prediction: An Empirical Study 2019 2nd International Conference on Advancements in Computational Sciences (ICACS) pp 1–7

[7] Al-Barrak M A and Al-Razgan M, 2016. Predicting students final GPA using decision trees: a case study Int. J. Inf. Educ. Technol. 6 528

[8] Zimmermann J, Brodersen K H, Heinimann H R and Buhmann J M, 2015. A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance J. Educ. Data Min. 7, pp. 151–76

[9] Elbattrawy A, Polyzou A, Ren Z, Sweeney M, Karypis G and Rangwala H 2016 Predicting student performance using personalized analytics Computer (Long. Beach. Calif). 49, pp.61–9

[10] Huang S and Fang N, 2013. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models Comput. Educ. 61, pp.133–45

[11] Kumar N,2014. Using Support Vector Machines Effectively

[12] Meier Y, Xu J, Atan O and der Schaar M, 2015. Predicting grades IEEE Trans. Signal Process. 64, pp.959–72

[13] Jian-Hui Y and Wei D, 2012. Prediction of gold price based on WT-SVR and EMD-SVR model 2012 Eighth International Conference on Computational Intelligence and Security pp 415–9

[14] Li D, Ju Y and Zou Q, 2016. Protein folds prediction with hierarchical structured SVM Curr. Proteomics 13, pp.79–85

[15] Kusumadewi S, 2004. Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Exel Link (Yogyakarta: Graha Ilmu)

[16] Ring M and Eskofier B M, 2016. An approximation of the Gaussian RBF kernel for efficient classification with SVMs Pattern Recognit. Lett. 84, pp.107–13

[17] Sudjana, 2005. Metode statistika Bandung: Tarsito

[18] Muflikhah L and Haryanto D J, 2018. High Performance of Polynomial Kernel at SVM Algorithm for Sentiment Analysis J. Inf. Technol. Comput. Sci. 3, pp.194–201