Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15 th October 202 5 . Vol. 5 . No. 1; e-ISSN: 2808-4519

Prediction of Academic Achievement of Vocational School Students Based on Tiktok Usage Patterns and Cognitive Styles: Multiple Linear Regression Model (Case Study: SMKS YPIS MAJU)

Khairi Syalwana Amin[1*], Suci Ramadani[2], I Gusti Prahmana[3]

[1,2,3] *STMIK Kaputama Binjai, Indonesia*
syalwanaamin@gmail.com[1*], suci.ramadani23@gmail.com[2], igustiprahmana4@gmail.com[3]

Abstract

This study aimed to predict vocational high school students' academic achievement by analyzing the influence of TikTok usage patterns and cognitive styles using a multiple linear regression model. The research was conducted at SMKS YPIS Maju Binjai with a sample of 100 students selected through purposive sampling. Data were obtained through questionnaires on TikTok usage patterns and cognitive styles, as well as students' academic records. TikTok usage (X1) was measured by frequency, duration, and its impact on study habits, while cognitive style (X2) was measured based on visual, verbal, and mixed learning preferences. Academic achievement (Y) was represented by students' average report card scores. The regression analysis produced the equation $\hat{Y} = 85.869 + (-0.3495X1) + (0.1993X2)$. The results showed that TikTok usage had a significant negative effect on academic achievement, whereas cognitive style had a significant positive effect. The model demonstrated good predictive accuracy with $R^2 = 0.392$, MAE = 1.77, MSE = 5.26, RMSE = 2.29, and MAPE = 2.16%. This study contributes by integrating social media usage patterns and cognitive factors to predict students' academic achievement and provides practical insights for educators in guiding students to balance social media use with academic learning.

Keywords: *Academic Achievement; Cognitive Style; Multiple Linear Regression; Prediction; Tiktok.*

1. Introduction

Academic achievement is an important indicator in assessing student success in school. Various factors can affect achievement, including family environment, motivation, learning strategies, and the use of digital technology. In recent years, social media has become an inseparable phenomenon from students' lives. TikTok, as one of the most popular social media platforms, has attracted attention due to its high usage among teenagers. However, overuse often negatively impacts students' concentration and academic achievement. [1]stated that the intensity of TikTok use can reduce learning focus, while [2]found a negative relationship between TikTok use and student learning outcomes.

In addition to social media factors, cognitive style also plays an important role in the learning process. Cognitive style describes the way an individual receives, processes, and remembers information. [3]) show that field dependent and field independent cognitive styles affect the mastery of the concept of Physics. Recent research by [4]also confirms that cognitive factors can be used as predictors of student academic achievement.

Several previous studies have examined the impact of social media on learning outcomes, as well as the influence of cognitive style on academic achievement. However, research that specifically combines TikTok usage patterns and cognitive styles in predicting students' academic achievement, especially at the vocational level, is still limited. Therefore, this study was conducted to fill the gap using a multiple linear regression model.

2. Litelature Riview

2.1. Machine Learning

Machine learning*)* is a branch of AI (Artificial intelligence) which can be used for the process of finding predictive values using popular algorithms including multiple linear regression [5]. Machine learning can make predictions, identify patterns, and even make decisions without being explicitly programmed due to the computer learning process from a given data set.

2.2. Predictions

According to [6] Predictions are obtained from the results of predicting or forecasting future events using relevant (historical) past data or information. The purpose of predicting is to get information on the greatest probability of what will happen in the future.

2.3. Academic Achievement

Academic achievement is the result or achievement obtained by a person in the field of education. Academic achievement can reflect the level of understanding, mastery of the material and achievements in a subject. According to [4] Teachers or lecturers assess student achievement by giving assessments to each student who takes their subjects, such as absenteeism, independent assignments, practicums, quizzes, Mid-Semester Exams and Final Semester Exams.

2.4. Tiktok

Tiktok is a social media platform that is popular among students. This short video-based application allows its users to create a video, share the video and watch short videos. According to [2] Tiktok may have both positive and negative effects on its users. Positive effects include sharing entertainment throughout your free time, working a lot, and learning more about the outside world. In addition, the negative effects of this type include often forgetting time, becoming addicted, paying less attention to health, and neglecting nearby places. Furthermore, the use of TikTok social media is one of the factors that can affect students' academic achievement

2.5. Cognitive Style

In general, cognitive style is a process of understanding and organizing in managing information. It is a person's natural preference or tendency to think, learn and also solve problems. One way to find out what intelligence students have is to pay attention to their cognitive style [7].

2.6. Multiple linear regression

The linear regression method is used to model the linear relationship between one or more independent variables (predictors) and one dependent variable (response). This model generates a straight-line equation that can be used to predict the value of a dependent variable after considering the value of an independent variable.

Multiple linear regression or commonly called MLR (regresi linier multivariat) is a statistical technique to predict the outcome of an answer variable, using a number of explanatory variables. The purpose of (MLR) is to model the linear relationship between the independent variables and the dependent variables to be analyzed [8].

$$Y = a + b_1X_1 + b_2X_2 + ... + b_kX_k \quad (1)$$

Information:

Y = Bound (Dependent) Variable
X_1 = First Independent Variable
X_2 = Second Independent Variable (Independent)
Xk = Next Free Variable if present (Independent)
$a + b_1 + b_2... + bk$ = Konstanta.

2.6.1 Prediction Testing

1. MAPE (Mean Absolute Percentage Error)

   Mean Absolute Percentage Error (MAPE) is a measure of relative determination used to calculate absolute error. According to [9] (MAP) is one of several methods used in evaluation to measure the accuracy or accuracy of a commonly used prediction result.

$$MAPE = \frac{1}{n}\sum \left| \frac{Y_{aktual} - Y_{prediksi}}{Y_{aktual}} \right| \otimes 100\% \quad (2)$$

2. MEA (Mean Absolute Error)

   MAE (Mean Absolute Error) is the average of the absolute value of the difference between the actual value and the value of the predicted result. MAE describes how large the average prediction error is without paying attention to the direction of the error (positive or negative). According to [10] MAE is the actual value and the prediction value whose value is absolutely positive.

$$MAE = \frac{1}{n}\sum \left| Y_{aktual} - Y_{prediksi} \right| \quad (3)$$

3. MSE (Mean Squared Error)

MSE (Mean Squared Error*)* is one of the evaluation metrics used to measure how large the average square error between the actual (real) value and the predicted value of a regression model. MSE is calculated by summing the square of the difference between each actual value and the predicted value, and then dividing it by the amount of data. According to [11] The average of the squared difference between the actual value and the predicted value in the dataset is represented by the Mean Squared Error (MSE).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{aktual} - y_{prediksi})^2 \quad (4)$$

4. RMSE (Root Mean Squared Error)

Root Mean Squared Error (RMSE) is an evaluation metric used to measure the average deviation, or error, between the actual value and the predicted value of a regression model. According to [11] RMSE is a continuation of the determination of MAE value.

$$RMSE = SQRT\left(\sum \frac{|y - y'|2}{n}\right) \quad (5)$$

5. R-Square (R^2)

R-Square (R^2) or commonly known as the coefficient of determination is a metric used to measure how much of the variation of the bound variable (Y) can be explained by the variable. free (X) in a regression model. data. According to [10] The determination coefficient shows how the free and non-free variables relate to each other.

$$R^2 = \frac{b_1(\sum x_1 y) + b_2(\sum x_2 y)}{\sum y^2} \quad (6)$$

3. Research Methodology

The research stage begins with data preprocessing, where questionnaire data on the Likert scale is changed and prepared to meet the analysis criteria. The research also uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as a guide to create predictive models., Here are the stages of the crips-dm pipeline used:

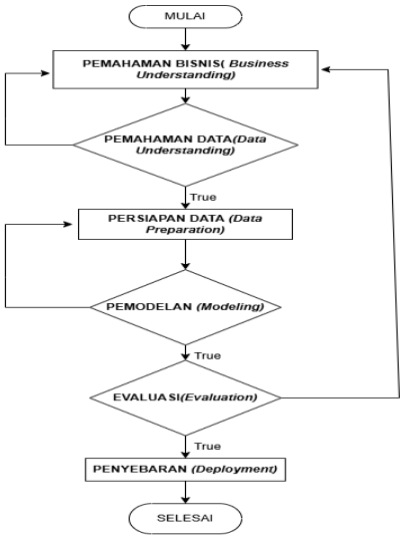

Fig. 1: Flowchart Crips-Dm

3.1. Types and Research Approaches

This study uses a quantitative type and uses an explanatory approach. The purpose of this methodology is to predict, test, and explain the causal relationship between the bound variable, i.e. the student's academic score (Y), and the independent variable, namely the pattern of Tiktok use (X1) and cognitive style (X2).

3.2. Population and Research Sample

The population in this study is students of SMKS YPIS MAJU Binjai who actively use Tiktok and have a record of academic grades. The purposive sampling method was used to collect 100 respondents.

3.3. Techniques and Instruments Data collection

The instrument used was a Likert scale questionnaire of 1-5 which was compiled based on theoretical indicators of Tiktok social media use and cognitive style. For data interpretation and analysis, all questions use positive redaction. Academic score data, which is a bound variable (Y) in this study, was obtained from the List of Value Sets (DKN) managed by the school. The final grade, which has gone through an official assessment process, shows the student's academic achievement. Meanwhile, data on TikTok usage patterns (X1) and cognitive styles (X2) were obtained through the distribution of questionnaires to students. The purpose of this questionnaire is to measure students' intensity and habits in using TikTok as well as their tendency to think and process data based on the concept of cognitive style.

3.4. Data Analysis Tools

A CRISP-DM-based multiple linear regression method with the help of Python in the Google colab was used to analyze the data results. The purpose of this method is to predict students' academic achievement based on pre-selected independent variables.

3.6. Input Data

3.6.1. Data retrieval

The initial stage begins with data collection. The questionnaire is distributed to obtain variables X1 and X2, and variable Y is obtained from the Value Set List, or DKN. One hundred respondents were collected from this data collection whose answers can be made for further analysis process.

Table 1. Questionnaire Sample DataSet

No	Name	Y	Q1	Q2	Q3	Q4	Q5	Q6	K1	K2	K3	K4
1	Savira Cantika	78.81	4	3	4	4	4	4	4	4	4	4
2	Febiviolandari	84.2	3	3	3	2	3	3	3	3	3	3
3	Fitri Saidah N	84.27	3	3	4	3	3	3	3	3	3	3
4	Shirin Sinta Bella	84.41667	3	3	4	3	3	3	3	3	3	3
5	Diva Elistia	86.72	3	3	4	3	3	3	3	3	4	3
6	Sri Ramadhani	85.125	3	3	2	2	2	3	4	4	4	3
7	Fahri Afrizal	82.833	3	3	2	4	4	3	4	4	4	4
8	Mesti Sassya Rukiani	88.5	3	3	3	2	2	3	3	4	4	3
9	Amanda Sabrina	84.125	4	3	4	3	3	3	4	4	5	5
…	….	…	…	…	…	…	…	…	…	…	…	…
100	Anisa Delia	79.247	5	4	4	2	4	4	5	5	4	2

The Y grade comes from the student's academic average in odd semesters in 2025. The Q1-Q6 scores indicate the student's tiktok usage patterns, and the K1-K4 scores indicate the students' cognitive style.

3.6.2. Data Preparation

The data is prepared for prediction processing using multiple linear regression algorithms. This is done to change the student's answer score. Variables X1 and X2, thus receiving the result of the following transformation:

Table 2 Data Transformation X1

	X1 raw data							Transformation Results
NO	Name	Q1	Q2	Q3	Q4	Q5	Q6	Total X1
1	Savira Cantika	4	3	4	4	4	4	23
2	Febiviolandari	3	3	3	2	3	3	17
3	Fitri Saidah N	3	3	4	3	3	3	19
4	Shirin Sinta Bella	3	3	4	3	3	3	19
5	Diva Elistia	3	3	4	3	3	3	19
6	Sri Ramadhani	3	3	2	2	2	3	15
7	Fahri Afrizal	3	3	2	4	4	3	19
8	Mesti Sassya	3	3	3	2	2	3	16
9	Amanda Sabrina	4	3	4	3	3	3	20

…	…	…	…	…	…	…	…	….
100	Anisa Delia	5	4	4	2	4	4	23

Table 3. Data Transformation X2

	X2 raw data					Transformation Results
NO	**Name**	**K1**	**K2**	**K3**	**K4**	**Total X2**
1	Savira Cantika	4	4	4	4	16
2	Febiviolandari	3	3	3	3	12
3	Fitri Saidah N	3	3	3	3	12
4	Shirin Sinta Bella	3	3	3	3	12
5	Diva Elistia	3	3	4	3	13
6	Sri Ramadhani	4	4	4	3	15
7	Fahri Afrizal	4	4	4	4	16
8	Mesti Sassya Rukiani	3	4	4	3	14
9	Amanda Sabrina	4	4	5	5	18
…	…	…	…	…	…	….
100	Anisa Delia	5	5	4	2	16

4. Implementation

The implementation stage is the implementation stage as well as the test. The implementation was carried out using the Python programming language with google collab media and data processing carried out manually.

Table 4. Research Dataset of Respondents (First 10 Samples Shown)

Y	Total X1	Total X2
78.81	23	16
84.2	17	12
84.27	19	12
84.41667	19	12
86.72	19	13
85.125	15	15
82.833	19	16
88.5	16	14
84.125	20	18
…	….	….
79.247	23	16

This table presents the overall data from 100 respondents by displaying the first 10 as a sample. The variable X1 represents TikTok usage patterns, and X2 represents cognitive style, both serving as independent variables. Meanwhile, Y refers to students' average academic scores, which act as the dependent variable. Subsequently, a multiple linear regression analysis was conducted to obtain the prediction results.

4.1. Results

a. Statistical Analysis Description Results

The following is a table of the results of the descriptive statistical analysis test of the data that has been tested using Python.

	X1	X2	Y
count	100.000000	100.000000	100.000000
mean	16.250000	13.920000	82.963152
std	4.759626	3.122677	2.958627
min	6.000000	4.000000	74.375000
25%	13.000000	12.000000	81.100250
50%	16.000000	14.000000	83.700000
75%	19.250000	16.000000	85.010417
max	28.000000	20.000000	89.111111

Fig. 2: Statistical Description of the Dataset

The results of the description also show that the pattern of TikTok use of SMK (X1) students averaged 16.25 from the assessment scale used, with a score distribution between 6 and 28. This shows that although there are some students who use TikTok very rarely, and there are also some who use it, it is very good. Students' cognitive styles (X2) showed differences in the way they thought and processed data, with an average of 13.92 and a score range of 4–20. The regression model built is expected to be able to stably predict academic achievement based on differences in TikTok usage patterns and students' cognitive styles. As a result, the academic score (Y) that was the focus of the prediction had an average of 82.96 and tended to be high, with a score range between 74.38 and 89.11.

b. Multiple Linear Regression Test Results

The results of multiple linear regression analysis yield the equation:

$$Y = a + b_1 X_1 + b_2 X_2$$

$$Y = 85.8690 + (-0.3495 X_1) + (0.1993 X_2) \quad (7)$$

The model shows that TikTok usage patterns (X1) have a negative impact on academic scores, while cognitive styles (X2) have a positive impact. Both variables were considered significant ($p < 0.05$), with a model strength of $R2 = 0.392$, suggesting that both variables could explain 39.2% of the variation in academic grades.

	X1	X2	Y	Prediksi_Y	Error
0	24	15	78.810000	80.47	-1.66
1	13	12	84.200000	83.72	0.48
2	19	12	84.270000	81.62	2.65
3	19	12	84.416667	81.62	2.80
4	10	14	86.720000	85.16	1.56
5	15	15	85.125000	83.62	1.50
6	16	14	82.833000	83.07	-0.24
7	7	16	88.500000	86.61	1.89
8	17	18	84.125000	83.51	0.61
9	8	16	87.540000	86.26	1.28

Fig.3: Prediction Results with 10 Samples

After testing by taking data of 10 samples, the results were obtained as shown in the image above.

c. Model Evaluation Results

```
== Metrik Evaluasi ==
Uji F           : F=31.3074, p=0.000000
R-squared       : 0.3923 (Adj 0.3798)
MAE/MSE/RMSE    : 1.7773 / 5.2664 / 2.2949
MAPE            : 2.16%

== Uji t (ringkas) ==
            coef          t             p
const  85.869000  60.062124  1.542968e-78
X1     -0.349544  -7.039686  2.765431e-10
X2      0.199299   2.633365  9.839596e-03
```

Fig. 4: Evaluation Metrics

The results of the evaluation of the F test model showed that the value of F = 31.3074 and the p-value = 0.000000 (< 0.05), which indicates that the regression model as a whole is significant and feasible to be used to predict academic grades. An R-

squared value of 0.3923 (Adj. R2 = 0.3798) indicates that 39.23% of the variation in students' academic grades is due to cognitive style and TikTok usage patterns, while the remaining 60.77% were influenced by other factors

The average prediction error values of MAE = 1.7773, MSE = 5.2664, and RMSE = 2.2949 indicate that the average difference is only about 2 points between the predicted value and the actual value. A MAPE value of 2.16% indicates a very low percentage error rate, so the model is considered accurate to use in predictions.

X1 factor (TikTok usage pattern): coefficient = -0.3495, t = -7.0397, p = 0.001 negatively and significantly impacted academic grades; the intensity of TikTok use is negatively correlated with students' academic grades. X2 factor (cognitive style): coefficient = 0.1993, t = 2.6334, p = 0.0098 had a positive and significant impact on academic grades. Students obtain higher academic grades if they have better cognitive, verbal and visual abilities.

4.2. Results Graph

Here is a graph of the histogram of the per-variable distribution:

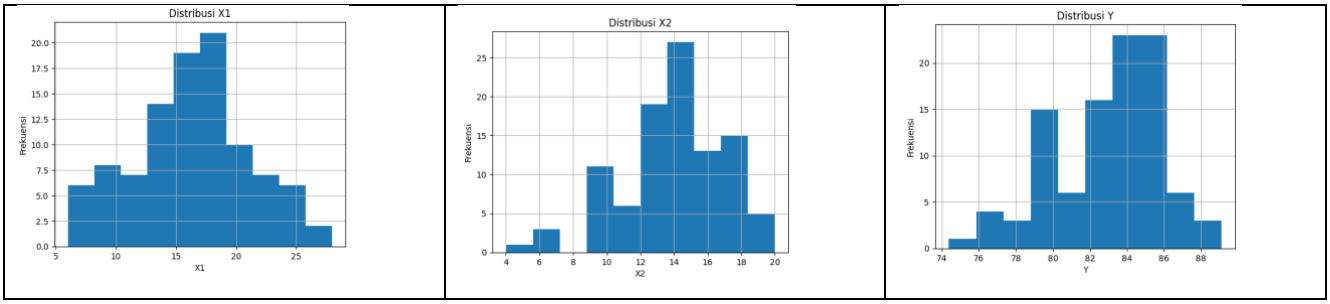

Fig. 5: Results of the Per-Variable Distribution Graph

The histogram shows that the majority of TikTok usage patterns (X1) are in the medium range (15–20), students' cognitive styles (X2) which include verbal and visual are mostly in the range of 12–16, and academic scores (Y) tend to be homogeneous, the majority are in the range of 82–86. This pattern shows that most students have moderate levels of TikTok use, relatively good cognitive style, and high academic scores.

Here is a graph of the independent per-variable correlation of the dependent variable:

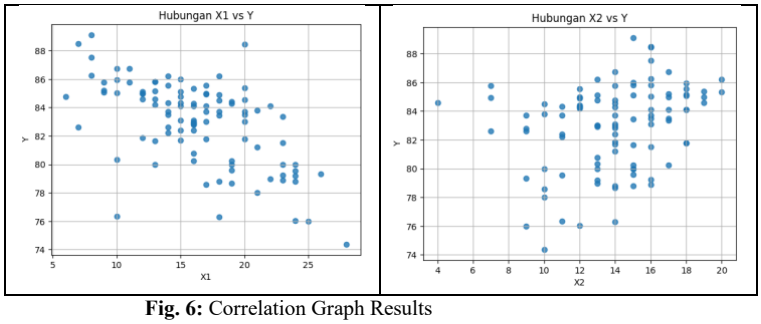

Fig. 6: Correlation Graph Results

Scatter plots show a negative relationship between TikTok usage patterns (X1) and academic scores (Y), where an increase in TikTok usage scores tends to be followed by a decrease in academic scores. On the other hand, the relationship between cognitive style (X2), which includes verbal and visual elements, and academic grades is positive, meaning that the higher a student's cognitive style score, the higher the academic grades they earn. This pattern is in accordance with the results of regression analysis, which shows that X1 has a significant negative impact and X2 has a significant positive impact on the academic achievement of vocational school students.

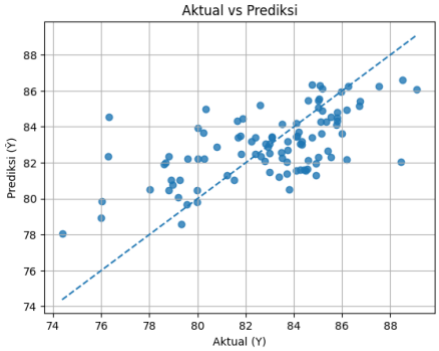

Fig. 7: Comparison Chart

Students' actual (actual) academic grades are compared to the predicted results of linear regression models in the Actual vs. Predicted graph. The model's prediction is getting closer to the actual value indicated by data points adjacent to the diagonal line. In accordance with the low MAPE value (2.16%) and RMSE of around 2.29 obtained earlier, the fairly tight distribution of data around this line suggests that the model has a good degree of accuracy. This suggests that the model has the ability to predict students' academic achievement with a very small error rate.

5. Conclusions and Suggestion

5.1. Conclusion

Based on the entire series of analyses that have been carried out by the researchers, the following conclusions are drawn:

1. The The Influence of TikTok Use Patterns (X1) on Academic Achievement. The results showed that indicators on the TikTok usage pattern questionnaire, such as the frequency of daily use, the duration of use of more than three hours daily, the tendency to use TikTok for entertainment rather than study, and its impact on study time and academic grades, overall negatively impacted the academic achievement of vocational school students. The higher the score on these elements, especially in the aspect of duration of use, has a negative impact on academic achievement. Nonetheless, there are some aspects that show a favorable outlook, namely the desire of students to use TikTok as a learning medium in schools, although the influence of the site has not yet become dominant.
2. The Influence of Cognitive Style (X2) on Academic Achievement. It is proven that students' cognitive styles have a positive impact on their academic achievement. These cognitive styles are measured through visual learning preferences (seeing pictures or diagrams), verbal (listening to the teacher's explanations), and the tendency to remember through sight rather than hearing. Students who Getting a high score on the cognitive style section of the questionnaire tends to get better academic scores. Students should have the ability to tailor their learning strategies to their cognitive strengths, which in turn will support their academic achievement.
3. The Contribution of the Second Variable to Academic Achievement. Simultaneously, the cognitive style and pattern of using TikTok have a major impact on the academic achievement of vocational school students. The variation in academic scores of 39.23% can be explained by the multiple linear regression model, which has high accuracy (MAPE 2.16% and RMSE 2.29). This suggests that cognitive styles can be beneficial for improving students' academic abilities, even though TikTok usage patterns have negative effects. Both of these styles can be used to predict academic achievement quite accurately.

These results provide students with a better understanding of the importance of controlling their use of TikTok so that it does not interfere with their studies; it is important for educators to adapt their teaching strategies to students' cognitive styles; and it is important for educational institutions to use social media as a basis for creating policies that maximize the use of social media in learning.

5.2. Suggestion

Based on the results of the conclusion of this study, the researcher provides several suggestions for future research:

1. Further research suggests adding other variables such as learning motivation, family support, or school environment that have the potential to affect students' academic achievement. This is done in order to increase the R2 value of the model and can explain the variety of academic achievement more broadly.
2. Further research is recommended to involve more than 1 school, both public and private, so that the results of the research can be generalized more widely.
3. The questionnaire can be expanded by adding details that learn about the type of content accessed on TikTok, the frequency of social interactions on the platform, and specific learning habits. By assessing a combination of kinesthetic, analytical, and reflective learning styles, instruments can be enriched for cognitive styles.

References

[1] Q. Aini, "Hubungan Media Sosial TikTok Terhadap Konsentrasi Belajar Siswa," *At-Taujih J. Bimbing. dan Konseling Islam*, vol. 2, no. 1, p. 3, 2023.

[2] S. R. Berliana, D. Mulyana, and C. Cahyono, "Pengaruh Penggunaan Media Sosial Tiktok Terhadap Hasil Belajar Peserta Didik Pada Mata Pelajaran Pendidikan Pancasila Dan Kewarganegaraan," *Melior J. Ris. Pendidik. dan Pembelajaran Indones.*, vol. 3, no. 2, pp. 39–48, 2023, doi: 10.56393/melior.v3i2.1720.

[3] M. G. Nugraha and S. Awalliyah, "Analisis Gaya Kognitif Field Dependent Dan Field Independent Terhadap Penguasaan Konsep Fisika Siswa Kelas Vii," no. February, pp. SNF2016-EER-71-SNF2016-EER-76, 2016, doi: 10.21009/0305010312.

[4] D. P. Wiguna, "Pemodelan Prediksi Mahasiswa Berprestasi Akademik Dengan Multi-Linear Regression (Studi Kasus : Prodi Sistem Informasi) Complex : Jurnal Multidisiplin Ilmu Nasional," vol. 2, pp. 53–58, 2025.

[5] M. Sholeh, S. Suraya, and D. Andayati, "Machine Linear untuk Analisis Regresi Linier Biaya Asuransi Kesehatan dengan Menggunakan Python Jupyter Notebook," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, p. 20, 2022, doi: 10.26418/jp.v8i1.48822.

[6] E. Hasibuan *et al.*, "Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web," *J. Ilm. Komputasi*, vol. 21, no. 4, pp. 595–602, 2022, doi: 10.32409/jikstik.21.4.3327.

[7] Astuti, S. B. Waluya, and M. Asikin, "Strategi Pembelajaran dalam Menghadapi Tantangan Era Revolusi Industri 4.0," *Semin. Nas. Pascasarj. 2019*, pp. 469–473, 2019, [Online]. Available: https://proceeding.unnes.ac.id/index.php/snpasca/article/view/327

[8] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 1, no. 2, pp. 140–147, 2020, doi: 10.38094/jastt1457.

[9] A. Anggara, K. Auliasari, and Y. Agus Pranoto, "Metode Regresi Linier Berganda Untuk Prediksi Omset Penyewaan Kamera Di Joe Kamera," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 852–858, 2023, doi: 10.36040/jati.v7i1.6158.

[10] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing," *Jambura J. Probab. Stat.*, vol. 4, no. 1, pp. 17–29, 2023, doi: 10.34312/jjps.v4i1.18948.

[11] H. Sulastri, G. S. Anwar, and E. N. F. Dewi, "Peramalan Stok Barang Percetakan dan ATK Menggunakan Single Moving Average," *J. Rekayasa Teknol. Inf.*, vol. 7, no. 1, p. 59, 2023, doi: 10.30872/jurti.v7i1.11876.