J-COSINE (Journal of Computer Science and Informatics Engineering)
Vol. 6, No. 2, December 2022 E-ISSN:2541-0806
Accredited Sinta-3 by RISTEKDIKTI Decree No. 28/E/KPT/2019 P-ISSN:2540-8895

Analysis and Prediction of GPA Based on UTBK Scores Using Linear Regression Algorithm

Ilham Firman Ashari*, Meida Cahyo Untoro, Sophia Nouriska,
Muhammad A.M Hikar Syahrial, and Makruf Alkarhi
Teknik Informatika, Institut Teknologi Sumatera
Jl. Terusan Ryacudu. Lampung Selatan. Indonesia
Email: firman.ashari@if.itera.ac.id, cahyo.untoro@if.itera.ac.id, sophia.119140027@student.itera.ac.id,
muhammad.119140031@student.itera.ac.id, muhammad.119140075@student.itera.ac.id

Corresponding Author

Abstract **Education is an important element that needs special attention, because education is an important element for the progress of a nation. Not only formal education starting from elementary school to high school, but it can also now be said that higher education has a central role in developing the capabilities of the Indonesian people. One way to get into state universities is with Ujian Tertulis Berbasis Komputer (UTBK). UTBK is set to be the main requirement to take part in the Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Students' abilities cannot necessarily be measured by the UTBK scores. Student achievement when entering tertiary institutions is determined by an assessment of the GPA (Cumulative Achievement Index). In this study the authors tried to find a relationship between the UTBK scores and student GPAs using the linear regression method. The results showed that the correlation between the UTBK and GPA scores was 0.670157 and the mean squared error was 0.0546.**

Key words**: SBMPTN, UTBK, Prediction, Linier Regression**

I. INTRODUCTION

Higher education is important and vital for the survival of the nation. With so many young people who are highly educated, quality human resources will be created [1]. In the process of continuing their education to higher education, a student must pass a university entrance exam at the national level called UTBK.

(UTBK) is an entrance test to college conducted by Lembaga Tes Masuk Perguruan Tinggi (LTMPT) as the organizer of standardized higher education tests in Indonesia [2]. LTMPT is responsible for organizing and managing the implementation of the Computer-Based Written Examination (UTBK) [3]. (UTBK) is set to be the main requirement to take part in (SBMPTN). The SBMPTN scores obtained by the examinees will be used to register at universities, then the universities will select scores that meet the requirements.

When in college, students are required to actively participate in class during teaching and learning activities. Student activity will support academic success because tasks can be completed effectively [4]. Student academic success is indicated by the ability of students to meet learning outcomes. Learning outcomes will be controlled by each university in order to produce quality graduates. The measure of learning success can be measured through the cumulative achievement index (GPA) obtained by students during the study period [5]. Students are declared to be able to pass the undergraduate program if they have successfully taken 144 credits of courses and achieved the target GPA greater than or equal to 2.00 [6].

With UTBK and GPA values that vary for each person. The question arises, "can the UTBK score be used to predict a student's GPA?". To make it easier to make predictions, a technology-based approach is needed using pre-existing data [7][8]. In making predictions there are several algorithms that can be used, namely SVM, Naive Bayes, C4.5, KNN, Random Forest, and Linear Regression [9][10].

Data is very important to support research success [11][12]. The data used in this study is data from students of informatics engineering class 2019. The linear regression method is used because it is based on research related to predictions by comparing the prediction algorithms SVM, Naive Bayes, C4.5, KNN, Random Forest, Linear Regression, showing that the linear regression algorithm is better in making predictions because it has the smallest RMSE value [10]. In this study the authors tried to find a relationship between the UTBK scores and student GPAs using the linear regression method. Then by making a training model the author tries to predict a student's GPA based on the UTBK scores obtained.

II. RELATED WORKS

Several studies conducted by previous researchers related to linear regression. The first research of [1] This study aims to determine the multiple linear regression model regarding the effect of GPA and study period on waiting time to get a job, and the most variable significant in influencing waiting time to get a job. The results of the study obtained a model to obtain the relationship between GPA and length of time. Research conducted by [13] related to the prediction of new students for the next 5 years with a simple linear regression method, the purpose of the research is to determine the ratio of lecturers to students so that lecture support facilities can be prepared.

Other research conducted by [10], The study was conducted to predict the increase in the number of Covid-19 cases in Indonesia, the results of the study showed that the linear regression algorithm was better with the smallest RMSE value than other prediction algorithms.

Another research conducted by Rahmana and Syarif [14] Regarding the prediction of sales and cash flow at the restaurant's point of sales, the results of linear regression from the study can be concluded that the increase in sales, capital, and profit are closely related to each other. Where if sales increase then the capital needed every month also increases so that it affects the profit earned every month. The prediction results of restaurant sales using linear regression using programming can get predictive results more quickly and accurately. By knowing the prediction of restaurant sales, you can take a strategy to maximize the resources you have.

Another study was conducted by Ayu, et al, who conducted a linear regression study for vegetable production [15]. Based on the results of the study, it can be concluded that: Linear Regression Method can be used to predict the amount of vegetable production at the Gorontalo Provincial Office. The results of testing the Linear Regression method with MAPE in predicting the amount of Vegetable Production based on 5 Types of Vegetables, namely the selected research results are shallots, cayenne pepper, kale, eggplant, and tomatoes. The error results obtained for red onion 35.013% with an accuracy rate of 64.987%, the error results obtained for cayenne pepper 15% with an accuracy rate of 85%, the error results obtained for kale 18.253% with an accuracy rate of 81.747%, the error results obtained for eggplant 85.638% with an accuracy rate of 14.362%, the error results obtained for tomatoes is 41.657% with an accuracy rate of 58.343%.

Research conducted by Alfanda, et al [16], is the application of multiple linear regression method to predict state losses based on corruption cases. Multiple linear regression learning in this study resulted in a regression model which resulted in a constant value of 284645.5891073216 and a coefficient value of -139837.38007863 and 363493.06049751. Then this study measures the performance of the linear regression model with the condition of data sharing 80% for training data and 20% for testing data. From the data sharing conditions, the RMSE value was 8447373,485 for training data and 9769609,026 for testing data. Meanwhile, the value of the coefficient of determination obtained a value of 0.579 for training data, which has a strong relationship between variables and 0.662 for testing data, which means that the relationship between variables is strong. By making predictions using the multiple linear regression method, it can provide information that helps the government in taking the right policies on corruption cases and minimizes and anticipates greater state losses for the following year.

Another research conducted by Trismawati and Astri [17]. This study aims to determine the multiple linear regression model regarding the effect of GPA and study period on waiting time to get a job, and the most significant variable in influencing waiting time to get a job. This study uses a quantitative descriptive approach. The subject of this research is Alumni of Mathematics Study Program FST UNDANA with a sample of 95 respondents. The data collection method used is a questionnaire. The analysis technique used is multiple linear regression analysis. The results of this study are multiple linear regression models for the effect of GPA and study period on waiting time to get the first job, namely Y=0.555-0.080X1+0.021X2, and there is no dominant variable in influencing waiting time to get a job because it is based on the results of partial hypothesis testing. The GPA and the study period for the Alumni of Mathematics Study Program FST UNDANA do not affect the waiting time to get the first job.

Subsequent research was carried out by Yumira, et al [18]. The purpose of this study was to determine the binary logistic regression model of the GPA of students of the Mathematics and Chemistry study program, FMIPA Unsrat Manado and to determine the factors that influence the GPA. The population in this study were active students of the Mathematics and Chemistry study program in the 2013/2014 academic year, amounting to 37 students and the 2014/2015 academic year totaling 49 students. The total population was 86 students, with the number of respondents who returned the research questionnaire as many as 80 respondents. Grade Point Average (GPA) as a response variable (Y) and gender (x1), major (x2), place of residence (x3), receiving a scholarship (x4), area of origin (x5), school origin (x6), occupation old age (x7), monthly cost of living (x8) as predictor variables in model fit.

Another study was conducted by Siti, et al., regarding the prediction of the number of hours of study on students' GPAs [19]. The purpose of this study was to determine the effect of the number of additional study hours on the GPA of students at the Bandung State Polytechnic (during the Covid-19 Pandemic). This study uses a quantitative method with multiple linear regression analysis, which begins with a literature study to determine aspects of additional study hours, cumulative achievement index, and distance learning during the Covid-19 pandemic. The results of the analysis of this study, which is based on multiple linear regression test, obtained an R square value of 0.124, with a sig F change of 0.060 > the value of = 0.05, then the decision is Ho is accepted and Ha is rejected. This means that the number of additional study hours does not significantly affect the GPA of students at the Bandung State Polytechnic (During the Covid19 Pandemic Period) because there are other factors that affect the student's GPA, namely self-interest, an environment that supports the learning atmosphere, and strategies. student learning itself.

Recent research conducted by Rosanti and M.yahya [20], This research was conducted at the Statistics Study Program, Faculty of Mathematics and Natural Sciences, Pattimura University using the multiple linear regression analysis method. It was found that the lecturers of the statistics study program, FMIPA Unpatti, had good

performance and the students' learning motivation was categorized as high. In addition, the Student Achievement Index (IPS) of the Unpatti FMIPA Statistics Study Program is in a very satisfactory category. As for the multiple regression analysis test, it was found that the variable of lecturer performance and student learning motivation of the Statistics Study Program FMIPA Unpatti could affect students' social studies with a Coefficient of Determination () of 0.685 (68.5%).

III. Research Method

At this stage, a discussion is carried out regarding the method used by the author when predicting student GPA based on UTBK scores with linear regression. The following is a description of the flow or work process carried out by the author, which can be seen in Figure 1.

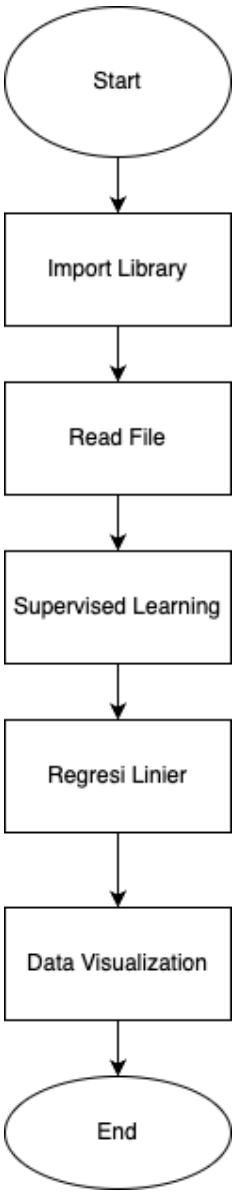

Fig 1. Research Flow

A. Importing Library

In this stage, the library installation will be used as shown in Figure 2 below.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

Fig 2. Importing Library

A.1. import numpy as np

NumPy (Numerical Python) is a library in python that has a focus on scientific computing [21][22]. This library helps us in linear algebraic operations, because the data processing we use uses the linear regression method.

A.2. import pandas as pd

Pandas (Python for Data Analysis) is a library in python that focuses on analyzing data such as data preparation, data manipulation, and data cleaning [21]. The operation on pandas that we use is a DataFrame, because the DataFrame operation is in accordance with the form of the data we are using, which is in the form of a data table containing columns and rows.

A.3. import matplotlib.pyplot as plt

Matplotlib is a library in python that focuses on data visualization such as graph plots [23]. Visualization in matplotlib is an image that presents a graph that has one or more axes. The horizontal axis (x) and the vertical axis (y), then the data presented has colors and glyphs that serve as markers, for example polygons, lines, and circles. Membaca File CSV

At this stage, the csv file that has been uploaded is read. After that, the file is called with the command pd.read_csv(filename) where the command is the default of the pandas library which aims to call a file with the extension .csv [23]. Then, do the df command to produce output when we run. Here is a picture where the author reads the csv file.

```
df = pd.read_csv('prediksi IPK.csv')
df
```

Fig 3. Reading CSV Files

B. Supervised Learning

Supervised Learning is one type of machine learning apart from Unsupervised Learning and Reinforcement Learning [24]. The algorithms belonging to the supervised learning type are Decision Tree, Naive Bayes, K-Nearest Neighbor, Super Vector Machine and Regression. Supervised learning uses datasets that have been trained to identify using machine learning to make predictions or

classifications. In this study, the author uses a regression algorithm in making predictions to find a correlation between the UTBK value and the GPA that will be obtained. Linear regression is also one of the data mining techniques where GPA predictions with data mining have been done several times with different parameters [25].

C. Linier Regression

There are several algorithms that can be used to find correlations between variables. The algorithms in question are Artificial Neural Network (ANN), fuzzy logic and linear regression [26]. Linear Regression is one of the methods in data mining or data mining, this method is usually used to show a relationship between one independent variable and has a straight line relationship with the dependent variable [27]. If there is only one independent variable in the regression model, it is called simple regression if more than one is called multiple linear regression [28].

Linear regression is a time series calculation that uses a quantitative method that uses time as the basis for its prediction. The basic equation for the simple linear regression method is in formula 1:

$$Y = a + bX \quad (1)$$

Description:
Y =Dependent Variable
a = Intercept
b = Variable coefficient X
X = Independent Variable

Furthermore, the following equations 2 and 3 can be used to find the intercept and variable coefficients:

$$a = \frac{(\sum X^2)(\sum Y) - (\sum XY)(\sum X)}{n(\sum X^2) - (\sum X)^2} \quad (2)$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \quad (3)$$

The observed variables are very likely to be influenced by other variables, for example, the data used in this study is the UTBK value and GPA predictions. For a certain GPA prediction there must be a magnitude of the UTBK value that affects it, and vice versa.

D. Data Visualitation

The last stage is data visualization. Data visualization is a display in the form of images that are presented digitally and non-digitally as one of the modern communication tools [29]. Data visualization is used so that the information and data displayed can be more easily understood by the reader so that it is possible to see the connection between the data based on the information provided [30]. At this stage, the author visualizes the data on the relationship between the GPA prediction data and the UTBK value. Where at this stage is a stage that is quite important because the selection of a data visualization method must match or match the data that has been processed.

IV. Result and Discussions

In conducting this research, the authors took 106 data samples, from the data of informatics students at the Institut Teknologi Sumatera. The data taken is the UTBK value data and the GPA of a student whose GPA will be determined in the future.

The first thing to do to start creating a prediction program is to import the library that will be used. The libraries used have been mentioned in the methods chapter. Then read the csv file containing the dataset used. The columns in the database are the UTBK and GPA values, totaling 106 data. This csv reading uses the pandas library with the *pd.read_csv* syntax.

After that, the writer calculates the statistical summary values such as mean, mode, median, and so on with the *df.describe()* syntax. The results obtained can be seen in Table I:

TABLE I. Summary of Dataset Statistics

Description	UTBK	IPK
Amount of Value	106.000000	106.000000
Average	840.000000	3.313585
Standard deviation	101.391274	0.281831
Minimum Value	634.000000	2.400000
25% percentile	764.000000	3.170000
50% percentile	837.000000	3.390000
75% percentile	934.000000	3.490000
Maximum Value	1000.000000	3.810000

Next, the correlation value between the two columns is calculated, namely the UTBK column and the GPA column with the **df.corr()** syntax. Correlation value is obtained by equation 4 below:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} \quad (4)$$

Description:
x : UTBK Value
$\bar{x}$: average value for utbk
y : nilai GPA value
$\bar{y}$: average value for GPA

By calculating the sum between the averages of each column and the sum of the squares of the two, the results are as in Table II.

TABLE II. Correlation Between Columns of Dataset

	UTBK	IPK
UTBK	1.000000	0.670157
IPK	0.670157	1.000000

The next step is to create a data visualization in the form of a scatter plot to see the distribution of data from the dataset used. The results obtained are as follows:

```
plt.figure(figsize=(10,8))
plt.scatter(df['UTBK'], df['IPK'], color='g',s=80)
plt.title('UTBK vs IPK')
plt.xlabel('UTBK')
plt.ylabel('IPK')
plt.show()
```

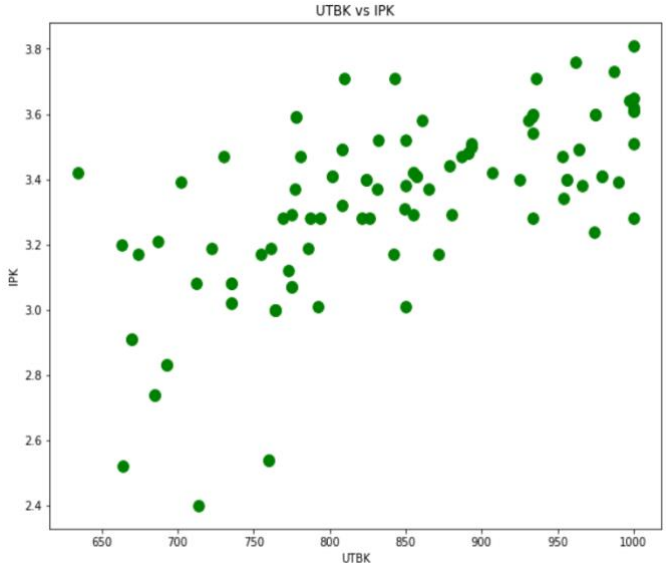

Fig 4. Initial Dataset Scatter Plot

Next, a training and testing period is carried out for the dataset used. The way it works is by dividing the dataset into two parts, namely the training set and the testing set, where the training data must be more than the testing data. Where in this case, 84 data are used for training sets, and 21 data are used for testing sets.

Training data is used to evaluate readiness and get the machine learning model fit from the dataset used. While Testing data is used to evaluate the results that have been obtained in the testing process. The illustrations for the depiction of the training set and testing set are as follows:

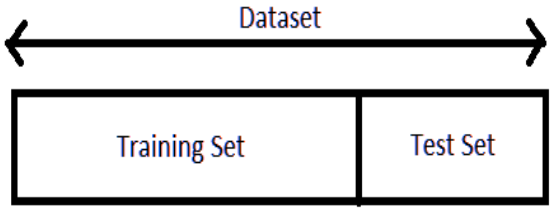

Fig 5. Illustration of Dataset Sharing for Training and Testing Sessions

In the process of training and testing this dataset, the Linear Regression algorithm which has been described in the previous chapter is used. The next thing to do after calculating the training set and testing set is to calculate the coefficient value and the intersection on the y-axis or often called the intercept. The results obtained can be seen in Table III:

TABLE III. COEFFICIENT VALUE AND INTERCEPT

coefficient	0.00172171
intercept	1.8677093164704834

The results of processing the dataset by conducting a training period using a linear regression algorithm represented by a scatter plot diagram can be seen in Figure 6.

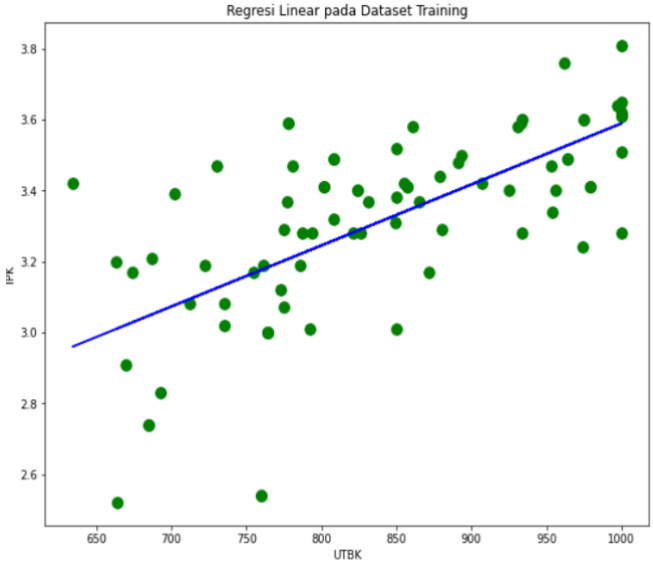

Fig 6. Representation of Linear Regression results in the Training Set

The results of processing the dataset during the testing period using a linear regression algorithm can be seen in Table IV.

TABLE IV. LINEAR REGRESSION RESULTS ON THE TESTING SET

UTBK Value	Actual Value of GPA	Predicted Value of GPA
891	3.48	3.40
735	3.02	3.13
693	2.83	3.06
987	3.73	3.57
670	2.91	3.02
832	3.52	3.30
769	3.28	3.19
855	3.29	3.34
855	3.42	3.34
775	3.07	3.20
956	3.40	3.51
966	3.38	3.53
714	2.40	3.10
843	3.71	3.32
842	3.17	3.32
893	3.51	3.41
934	3.54	3.48
936	3.71	3.48
887	3.47	3.39
990	3.39	3.57
735	3.08	3.13
810	3.71	3.26

The above results can also be represented in the form of a scatter plot diagram with a linear regression line. The representation can be seen in Figure 7.

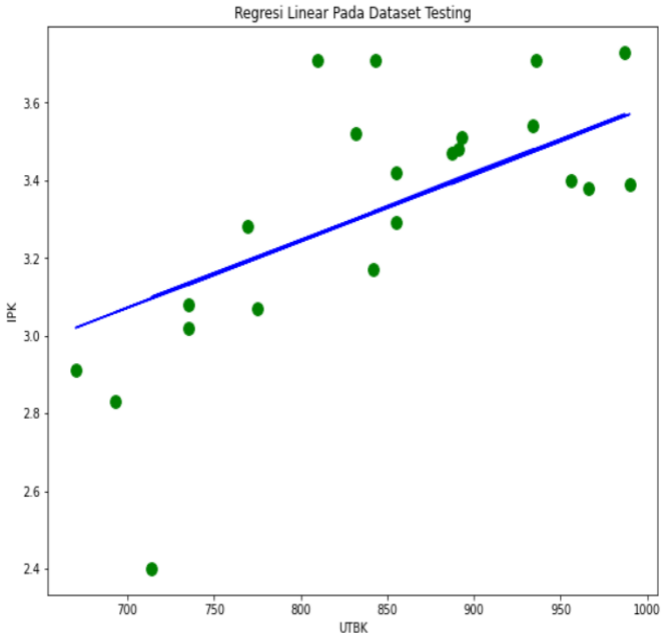

Fig 7. Representation of Linear Regression results on the Testing Set

The smaller the MAE (Mean Absolute Error) value, the higher the level of accuracy. The MSE and MAE equations can be seen in equation 6. The program code for calculating MSE and MAE can be seen in Figure 8.

```
from sklearn.metrics import mean_squared_error, mean_absolute_error
print('Mean Squared Error: ', mean_squared_error(y_test, y_pred))
print('Mean Absolute Error: ', mean_absolute_error(y_test, y_pred))
```

Fig 8. Code Program For Count MSE and MAE

$$MSE = \sum \frac{(y'-y)^2}{n} \quad (5)$$

Y ' = Predicted Value
Y = Actual Value
n = The number of data

$$MAE = \sum \frac{(|y'-y|^2}{n} * 100\ \% \quad (6)$$

Y ' = Predicted Value
Y = Actual Value
n = The number of data

The value of the difference between the target output value and the actual output value is 0.0546 and the absolute difference between the predicted and actual values is 0.1778. Can be represented in table 5 below:

TABLE 5. MSE AND MAPE VALUES

Mean Squared Error	0.05458636363636365
Mean Absolute Error	0.1777272727272727

V. CONCLUSIONS

Based on the research that has been done, it can be concluded and it is known that the correlation value between the columns of UTBK and GPA values in the dataset used is 0.670157. The correlation shows that the higher the UTBK value, the greater the GPA that will be obtained. The coefficient value is 0.00172171 and the intercept is 1.8677093164704834. Linear regression carried out at the training and testing stages showed few outliers so that it can be said to meet the target. The difference between the initial value and the prediction is obtained with a mean squared error of 0.0546 and a mean absolute error of 0.1778 where the error value is relatively small.

Suggestions for further research is to use more complete data with all samples based on the entire study program at the Sumatra Institute of Technology and make applications to test the model.

REFERENCES

[1] T. P. Kiri and A. Atti, "Pengaruh IPK Dan Masa Studi Terhadap Waktu Tunggu Mendapatkan Pekerjaan," *J. Difer.*, vol. 03, no. 01, pp. 1–12, 2021, [Online]. Available: http://ejurnal.undana.ac.id/index.php/JD/article/view/3998.

[2] H. Ridwan, I. Darmawati, and M. N. Rahmawati, "Implementasi Kebijakan Physical Distancing dalam Pelaksanaan UTBK SBMPTN di Perguruan Tinggi," *J. Kebijak. Kesehat. Indones. Jkki*, vol. 09, no. 04, pp. 197–201, 2020, doi: 10.22146/jkki.59689.

[3] M. Mukminina and Z. Abidin, "Coping Kecemasan Siswa SMA dalam Menghadapi Ujian Tulis Berbasis Komputer (UTBK) Tahun 2019," *J. Al-AZHAR Indones. SERI Hum.*, vol. 5, no. 3, p. 110, 2020, doi: 10.36722/sh.v5i3.384.

[4] S. Imaslihkah, M. Ratna, and V. Ratnasari, "Analisis regresi logistik ordinal terhadap faktor-faktor yang mempengaruhi predikat kelulusan mahasiswa S1 di ITS Surabaya," *J. Sains dan Seni Pomits*, vol. 2, no. 2, pp. 177–182, 2013.

[5] L. Safitri Linawati, Siti Nurdiani, Kartika Handayani, "Prediksi Prestasi Akademik Mahasiswa Dengan Menggunakan Algoritma Random Forest dan C4.5," *2JURNAL KHATULISTIWA Inform.*, vol. VIII, no. 1, pp. 6–13, 2020, [Online]. Available: https://www.academia.edu/10982938/Prediksi_Prestasi_Mahasiswa_Dengan_Metode_Jaringan_Saraf_Tiruan.

[6] Z. Hodsay, "Perbedaan Indeks Prestasi Kumulatif (IPK) Mahasiswa Antara Gaya Belajar Visual, Auditorial dan Kinestetik Pada Program Studi Pendidikan Akuntansi

FKIP Universitas PGRI Palembang," *J. Profit*, vol. 3, no. 1, pp. 73–86, 2016.

[7] I. F. Ashari, "Implementation of Cyber-Physical-Social System Based on Service Oriented Architecture in Smart Tourism," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 66–73, 2020, doi: 10.30871/jaic.v4i1.2077.

[8] I. F. Ashari, R. Banjarnahor, and D. R. Farida, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," vol. 6, no. 1, pp. 7–15, 2022.

[9] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[10] dewi wulan Sari and mitha maharani Wahyudi, "Analisis Dan Perbandingan Algoritma Prediksi Dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus Covid-19 Di Indonesia Dengan Metodologi CRISP-DM," no. June, 2021.

[11] I. F. Ashari and V. Adhelia, "Expert System and IoT for Diagnose of Feline Panleukopenia Virus Using Certainty Factor," *Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 451–462, 2022, doi: 10.30812/matrik.v21i2.1517.

[12] I. F. Ashari, M. Alfarizi, M. N. K, and M. A. H, "Vulnerability Analysis and Proven On The neonime . co Website Using OWASP ZAP 4 and XSpear," *J. Teknol. Komput. dan Sist. Inf.*, vol. 5, no. 2, pp. 75–81, 2022.

[13] N. Almumtazah, N. Azizah, Y. L. Putri, and D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 1, pp. 31–40, 2021, doi: 10.22487/2540766x.2021.v18.i1.15465.

[14] R. D. Shaputra and S. Hidayat, "Implementasi Regresi Linier Untuk Prediksi Penjualan Dan Cashflow Pada Aplikasi Point Of Sales Kafe Xyz," *Automata*, vol. 2, no. 1, pp. 1–7, 2021.

[15] A. A. Basahona, I. Rezqiwati, and A. husna N, "Penerapan Metode Linier Regresi Untuk Prediksi Produksi Sayur-Sayuran," *IC Tech*, vol. XIV, no. 2, pp. 50–53, 2019.

[16] A. N. Maharadja, I. Maulana, and B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021, doi: 10.30871/jaic.v5i1.3184.

[17] T. P. Kiri and A. Atti, "Pengaruh IPK Dan Masa Studi Terhadap Waktu Tunggu Mendapatkan Pekerjaan," *J. Differ.*, vol. 03, no. April, pp. 1–12, 2021.

[18] Y. A. Tampil, H. Komalig, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *JdC*, vol. 6, no. 2, pp. 56–62, 2021, doi: 10.35799/dc.6.2.2017.17023.

[19] S. Karlina, R. S. Hayati, C. Danari, Putri, and N. Nuryati, "Pengaruh jumlah jam belajar tambahan terhadap indeks prestasi kumulatif mahasiswa: studi kasus di Politeknik Negeri Bandung (selama masa pandemi covid-19)," *Pros. 12th Ind. Res. Work. Natl. Semin.*, pp. 1574–1579, 2021, [Online]. Available: https://jurnal.polban.ac.id/ojs-3.1.2/proceeding/article/view/2987/2313.

[20] R. Laing and M. Yahya Matdoan, "Penerapan Regresi Linier Berganda Untuk Mengetahui Pengaruh Kinerja Dosen Dan Motivasi Belajar Terhadap Ips Mahasiswa Dalam Pembelajaran Daring," *J. Mat. Stat. dan Terap.*, vol. 1, no. 02, pp. 68–78, 2021, [Online]. Available: https://ojs3.unpatti.ac.id/index.php/parameter.

[21] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.

[22] I. F. Ashari, A. G. Manalu, and R. Setiawan, "Analysis of Security Guard Scheduling System Using Genetic Algorithm and Tournament Selection (Case Study : Institut Teknologi Sumatera)," vol. 5, no. 2, pp. 202–207, 2021.

[23] F. T. Industri *et al.*, "Image Processing / Augmentation With Scikit-Image," vol. 13, 2022.

[24] N. L. A. Puput Santaso, Heri Abijono, "Algotitma Supervised Learning dan Unsupervised Learning Dalam Pengolahan Data," *J. Teknol. Terap.*, vol. 4, no. 2, 2021.

[25] D. R. Anita Desiani, Sugandi Yahdin, "Prediksi Tingkat Indeks Prestasi Akumulatif Akademik Mahasiswa Dengan Menggunakan Teknik Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1237–1244, 2020.

[26] T. N. Padilah and R. I. Adam, "Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang," *FIBONACCI J. Pendidik. Mat. dan Mat.*, vol. 5, no. 2, p. 117, 2019, doi: 10.24853/fbc.5.2.117-128.

[27] H. BASRI, "Pemodelan Regresi Berganda Untuk Data Dalam Studi Kecerdasan Emosional," *Didakt. J. Kependidikan*, vol. 12, no. 2, pp. 103–116, 2019, doi: 10.30863/didaktika.v12i2.179.

[28] T. Indarwati, T. Irawati, and E. Rimawati, "Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone," *J. Teknol. Inf. dan Komun.*, vol. 6, no. 2, pp. 2–7, 2019, doi: 10.30646/tikomsin.v6i2.369.

[29] A. M. B. Nadiar Ahmad Syaripul, "Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi DKI Jakarta : Topik Ekonomi dan Keuangan Daerah," *J. Sist. Inf.*, vol. 12, no. 2, pp. 82–89, 2016.

[30] Y. A. Ghani, "Pengembangan Sarana Prasarana Destinasi Pariwisata Berbasis Budaya di Jawa Barat," *J. Pariwisata*, vol. IV, no. 1, pp. 22–31, 2017.