E-ISSN : 2963-4946
Vol. 2 No. March 6, 2024

Asian Journal of Social and Humanities
https://ajosh.org/

Assessment of The Level of Student Understanding In The Distance Learning Process Using Marchine Learning

Adilah Widiasti[1], Agung Mulyo Widodo[2*], Gerry Firmansyah[3], Budi Tjahjono[4]
Universitas Esa Unggul, Indonesia
E-mail: adilahwidiasti66@gmail.com[1*], agung.mulyo@esaunggul.ac.id[2], gerry@esaunggul.ac.id[3], budi.tjahyono@esaunggul.ac.id[4]

*Correspondence: agung.mulyo@esaunggul.ac.id

KEYWORDS
Artificial Intelligence; Marchine Learning; Classification Techniques; Clustering Technique; Logistic Regression; K-Means

ABSTRACT
As technology develops, data mining technology is created and used to analyse students' level of understanding. This analysis is conducted to group students according to their ability to understand and master the subject matter. This research can provide guidance and insight for educators, as well as artificial intelligence, marchine learning, association techniques, and classification techniques. Researchers and policymakers are working to optimise learning and improve the quality of student understanding. This study aims to analyse student understanding in simple and structured terms. Using the Marchine learning method to analyse the level of student understanding has the potential to impact the quality of education significantly. In addition, marchine learning categories are qualified to be applied to the concept of data mining. The data mining techniques used are association and classification. Association techniques are used to determine the pattern of distance student learning. The following process of classification techniques is used to determine the variables to be used in this study using the Logistic Regression model where data that have been classified are grouped or clustered using the K-Means algorithm into three, namely the level of understanding is excellent, sound, and lacking, based on student activity, assignment scores, quiz scores, UTS scores, and UAS scores.

Introduction

Over time, education in Indonesia is currently facing the biggest problem, namely the COVID-19 pandemic, which has resulted in a limited learning process from learning that is usually done face-to-face must be expert in distance learning (Azzahra, 2020). The absence of a face-to-face learning process directly at school or university will make it difficult for teachers or lecturers to deliver learning materials (Anars et al., 2018). Not only teachers or lecturers, students or students also experience the same difficulty in understanding the material. Many factors influence this, such as the lack of facilities owned by students or students (Herliandry et al., 2020). This is due to the development of the virus that spreads rapidly throughout the world, especially in Indonesia, which impacts education. This affects changes in policies that have been implemented. New educational policies change learning, which is usually carried out face-to-face, but due to the COVID-19 pandemic, learning is finally carried out through distance learning. Technology is currently considered a substitute for the position of teachers or lecturers in learning (Al-Azawei & Alowayr, 2020).

Education has undergone significant transformation and technological developments, especially in distance learning. Many educational institutions, including colleges and universities, have been forced to switch to distance learning. Distance learning utilises information and communication technology to deliver learning materials to students without physical presence in the classroom and is carried out remotely. I will schedule some time for us to connect. Virtually. When the distance learning process only sometimes runs smoothly, some obstacles can occur (Halim & Prasetyo, 2018). One of the problems faced is the need for more to measure students' level of understanding and skills objectively. The material presented needs to be more explicit, and students also need help understanding what lecturers convey. There needs to be more student discipline in collecting assignments. If students need help understanding the previous material, it will be easy to understand the following material (Yulianti et al., 2023). So, students need help understanding the class lessons. For lecturers who teach courses in class, it is a problem to provide material because of uneven student understanding; this will cause learning goals that the semester study plan cannot achieve. The existence of obstacles in the learning process causes several changes in students, ultimately affecting the decline in learning achievement (Rochmawati et al., 2023).

In overcoming this challenge, Marchine learning technology is part of artificial intelligence (Artificial Intelligence). The role of Marchine learning and Artificial Intelligence enables faster, more accurate, and objective analysis of assessment study data. Predictions and classifications can be made based on the student's understanding of the data collected using artificial intelligence algorithms and marchine learning models. This allows the identification of patterns, factors, and variables that influence students' understanding more systematically and deeply (Wahyudi et al., 2023). By combining marchine learning techniques with assessment studies, educators and researchers can better understand learning effectiveness, detect students' difficulties, and develop more adaptive and personalised learning strategies. In addition, machine learning categories are qualified to be applied to the concept of data mining.

Data mining techniques are classification and clustering. Classification techniques are used to determine the variables in this study, using a prediction algorithm for student learning performance in the distance learning process, namely the Logistic Regression algorithm (Sidik et al., 2020). Logistic Regression algorithm is a popular classification algorithm often used in marchine learning and data mining. Logistic Regression algorithms are often used in binary prediction or classification, where the target or dependent variable is a binary category (e.g., yes/no, success/failure, 1/0). Logistic regression algorithms are used to predict students' level of understanding because the logistic regression results can be interpreted easily. Regression coefficients can provide information about how each independent variable contributes to the probability of a target category event. Logistic regression tends to be more computationally efficient than more complex classification models. This makes it suitable for datasets with large sizes or with a large number of variables. Logistic regression can handle category variables well. In the context of predicting the level of understanding, students produce probabilities for each target category, which can provide more information about the level of confidence of the model in making predictions, variables such as learning modules, discussion forums, assignment/quiz grades, UTS scores, and UAS scores. The level of accuracy varies, but the highest level of accuracy is used to analyse the level of student understanding in virtual learning using the Logistic Regression algorithm.

The clustering technique used is the K-Means algorithm. K-means can be used to group students into groups based on their level of understanding. This allows lecturers to understand how knowledge of the material is spread among students and identify groups that require additional attention by knowing how students understand the learning material by being divided into three categories: excellent, pleasing and not good.

Based on the background described in this study, in the last five years, classification research has used marchine learning (appendix 1), especially to analyse learning. Several recent studies highlight the role of artificial intelligence in improving the effectiveness of assessing students' level of understanding of distance learning. According to (Kim et al., 2019), artificial intelligence technology in distance learning can help monitor student learning progress in real time and provide recommendations tailored to individual needs. Similarly, according to research by (Huang et al., 2021), Artificial Intelligence has opened up new opportunities for developing an adaptive evaluation system for students' abilities and learning preferences. According to a study by (Kim et al., 2019), using marchine learning techniques in online data analysis can help identify different learning patterns and optimise learning strategies accordingly. Research (Bubb & Jones, 2020) shows that logistic regression effectively and accurately predicts student success rates in online courses. Meanwhile, research (Merliana, 2015) found that clustering k-means can be used to identify different learning patterns among college students in a distance learning environment.

Based on the background stated above, the objectives of this Thesis Report are as follows:

1. Evaluate the effectiveness of using marchine learning in analysing the level of student understanding in learning based on assessment study data.
2. Testing the accuracy and objectivity of AI methods in predicting the level of student understanding based on test results and student attributes.
3. Identify factors contributing to students' understanding of learning, including student attributes and critical questions in assessment studies.
4. Explain the implications of research results in an educational context and provide practical recommendations for educators to improve student understanding in a virtual learning environment

Research Methods

Object of Research

The object of this research is XYZ University data; researchers conducted research at the faculty of computer science, majoring in informatics engineering with courses during the COVID-19 pandemic, including Algorithms and programming, Data Structures, Database Systems, Artificial Intelligence, and Software Engineering. This data is for 2020-2021.

Research Methods

The research methods used to achieve the objectives formulated in the study are as follows:

1. Data Collection Techniques

Observation

At this stage, direct observation was carried out, where researchers came directly to the location to see field conditions and the learning process at XYZ University. Researchers conducted research at the faculty of computer science, majoring in informatics engineering with courses during the COVID-19 pandemic, including Algorithms and programming, Data Structures, Database Systems, Artificial Intelligence, and Software Engineering.

Research Schedule

Research schedule, which includes preparation, implementation and reporting of research results. The research will be carried out within one semester from July 2020 to June 2021 (Appendix 2).

2. Literature Review

Literature studies use sources such as National or International Journals with a minimum range of 5 years and a maximum of 10 years. Apart from journals, other sources are books and articles. This literature study helps find the theoretical basis, knowledge, and information about the internal and external environment.

Making Observations

At this stage, direct observation is conducted, where researchers come directly to the location to see field conditions and how the learning process is virtual. The results of observations can be seen from the ongoing business processes described using BPMN.

Identify the Problem

At this stage, researchers identify problems based on data obtained from observation, questionnaires, and literature studies.

Data Collection

The data collected is secondary data from XYZ University, at the computer science faculty, majoring in informatics engineering with courses during the COVID-19 pandemic, including Algorithms and programming, Data Structures, Database Systems, Artificial Intelligence, and Software Engineering. This data is for 2020-2021.

Data Processing

Data processing is done with the help of Python. This is because using Python makes it possible to run machine learning modules to simplify and speed up data processing.

Results and Discussions

The data collected is secondary data from XYZ University, at the computer science faculty, majoring in informatics engineering with courses during the COVID-19 pandemic, including Algorithms and programming, Data Structures, Database Systems, Artificial Intelligence, and Software Engineering. This data is for 2020-2021. This data may include attributes such as opening the homepage, lecture modules, frequency of participation in discussions, assignment scores, UTS scores and UAS scores. The process is carried out by processing the data using Python, which consists of Pandas, NumPy, Matplotlib, and Scikit-Learn. Identify features or variables that may affect the level of student understanding. This can include attributes such as duration of study per week, number of completed assignments, previous exam scores, attendance rate in virtual classes, and participation in online discussion forums.

Model Training

At this stage, the data that has been collected has attributes including Grades, Homepage, Lecture Modules, Forums, Assignments, Quizzes, UTS, UAS and Labels. The collected data as many as 1428 data were collected and saved into Excel files with .xlsx format. There is no limit to the amount of data retrieved, but 1428 data is expected to represent the results of general student opinion. The data collected in Excel format can be seen in the figure below:

B1427 | 0

	A	B	C	D	E	F	G	H	I	J
1	NILLAI	Homepage	Modul Perkuliahan	Forum	Tugas	Kuis	UTS	UAS	Label	
2	80	1	1	0	1	0	0	1	1	
3	75	1	0	1	1	1	1	1	1	
4	88	1	1	1	1	1	1	1	1	
5	86	1	1	0	0	0	1	1	1	
6	84	1	1	0	1	0	0	1	1	
7	82	1	0	1	1	1	1	1	1	
8	77	1	1	1	1	1	1	1	1	
9	78	1	1	0	0	0	1	1	1	
10	83	1	1	0	1	0	0	1	1	
11	85	1	0	1	1	1	1	0	1	
12	87	1	1	1	1	1	1	0	1	
13	76	1	1	0	0	0	1	0	1	
14	79	1	1	0	1	0	0	0	1	
15	81	1	0	1	1	1	1	0	1	
16	89	1	1	1	1	1	1	0	1	
17	90	1	1	0	0	0	1	0	1	
18	93	1	1	0	1	0	0	1	1	
19	92	1	0	1	1	1	1	1	1	
20	91	1	1	1	1	1	1	1	1	

	Homepage	Modul Perkuliahan	Forum	Tugas	Kuis	UTS	UAS	Label
count	1427.000000	1427.000000	1427.000000	1427.000000	1427.000000	1427.000000	1427.000000	1427.000000
mean	0.529082	0.529082	0.529082	0.508059	0.361598	0.442186	0.466713	0.463910
std	0.499329	0.499329	0.499329	0.500110	0.480632	0.496820	0.499066	0.498871
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000
75%	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

The data that has been collected will be separated into two sets, namely the training set and the testing/validation set. The training set will be used to train the model, while the test set will be used to test the performance of the model that has been introduced. The data is separated into two parts: the training and testing sets. Divide the data into 70% training and 20% testing.

```
x_test
```

	NILLAI	Homepage	Modul Perkuliahan	Forum	Tugas	Kuis	UTS	UAS
226	68	1	0	1	1	1	1	1
264	70	0	1	0	1	0	1	0
490	78	1	1	0	1	0	0	1
534	85	1	1	1	1	1	1	1
1053	55	0	1	1	1	1	1	1
...	...	...	...	...	...	...	...	...
321	77	0	1	0	1	0	1	0
817	95	0	1	0	1	0	1	0
348	79	0	1	0	1	0	1	0
391	81	1	0	1	1	1	1	0
139	68	1	0	1	1	1	1	1

357 rows × 8 columns

```
x_train
```

	NILLAI	Homepage	Modul Perkuliahan	Forum	Tugas	Kuis	UTS	UAS
836	92	1	1	1	1	1	1	1
1317	89	0	1	0	1	0	0	0
387	85	1	1	0	1	0	0	1
936	51	1	1	1	1	1	1	1
1275	54	0	1	0	0	0	1	0
...	...	...	...	...	...	...	...	...
837	91	1	1	0	0	0	1	1
987	88	0	1	0	1	0	1	0
755	94	0	1	0	1	0	1	0
1096	76	0	1	0	1	0	0	1
1254	89	0	1	1	1	1	1	1

1070 rows × 8 columns

Data Preprocessing

The data obtained from the observations at the previous stage cannot be used because it still has an unstructured sentence form, so it is necessary to preprocess the data. Preprocessing removes noise and converts the data obtained to suit the needs by enlarging or reducing the data value.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1427 entries, 0 to 1426
Data columns (total 8 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Homepage            1427 non-null   int64
 1   Modul Perkuliahan   1427 non-null   int64
 2   Forum               1427 non-null   int64
 3   Tugas               1427 non-null   int64
 4   Kuis                1427 non-null   int64
 5   UTS                 1427 non-null   int64
 6   UAS                 1427 non-null   int64
 7   Label               1427 non-null   int64
dtypes: int64(8)
memory usage: 89.3 KB
```

```
Homepage             1427
Modul Perkuliahan    1427
Forum                1427
Tugas                1427
Kuis                 1427
UTS                  1427
UAS                  1427
Label                1427
dtype: int64
```

At this point, null and unnecessary data is cleaned before classification and provides information about the DataFrame (pdf), including the number of rows and columns, the data type of each column, and the number of non-null values (no missing values).

Labelling

I separate features and targets from the Data Frame (pdf) when creating a classification model.

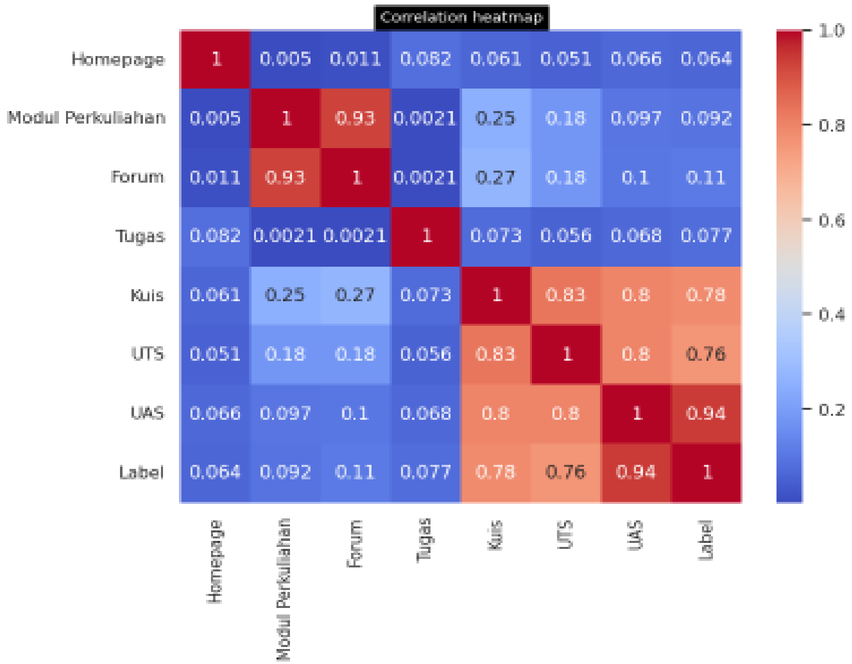
```
[ ]  x = df.drop(["Label"], axis = 1)
     y = df["Label"]
```

At this stage, the data is classified by separating features and targets from the existing raw data—a way to prepare data for logistic regression. Typically, in statistical modelling such as logistic regression, we need two datasets: one dataset contains the independent feature or variable (usually called X), and one dataset contains the target or dependent variable you want to predict (traditionally called y).

Correlation Table

At this stage, the heatmap displays the correlation between the features in the dataframe. The relationship between features can provide insight into correlated or interrelated features. To interpret the strength of the relationship between two variables is done by looking at the correlation coefficient number of the calculation results using the following criteria:

1. If the correlation efficiency number shows 0, then the two variables have no relationship
2. If the correlation efficiency number is close to 1, then the two variables have a stronger relationship
3. If the correlation efficiency number is close to 0, then the two variables have a weaker relationship
4. If the correlation coefficient number equals 1, both variables have a positive perfect linear relationship.
5. If the correlation coefficient number equals -1, both variables have a negative, perfectly linear relationship.



Tabel Korelasi:

	Homepage	Modul Perkuliahan	Forum	Tugas	Kuis	UTS	UAS	Label
Homepage	1.000000	0.005014	0.010644	0.082246	0.060650	0.051404	0.065991	0.063598
Modul Perkuliahan	0.005014	1.000000	0.926872	0.002147	0.251275	0.178510	0.097456	0.092191
Forum	0.010644	0.926872	1.000000	0.002147	0.268807	0.181337	0.103084	0.109082
Tugas	0.082246	0.002147	0.002147	1.000000	0.073454	0.055653	0.068079	0.076533
Kuis	0.060650	0.251275	0.268807	0.073454	1.000000	0.830610	0.798643	0.782713
UTS	0.051404	0.178510	0.181337	0.055653	0.830610	1.000000	0.799002	0.759048
UAS	0.065991	0.097456	0.103084	0.068079	0.798643	0.799002	1.000000	0.938049
Label	0.063598	0.092191	0.109082	0.076533	0.782713	0.759048	0.938049	1.000000

Explanation of the results of the correlation table between variables includes:

a. Homepage: Has a very weak correlation with Forums, Assignments, Quizzes, UTS, UAS, and Labels. This shows that the homepage does not have a significant influence on the variables Forum, Assignments, Quizzes, UTS, UAS, and Labels.
b. Lecture Module: Has a strong and positive correlation with Forums, Assignments, Quizzes, UTS, UAS, and Labels. This shows that the more lecture modules available, the higher the level of student activity and participation in forums, assignment work, quizzes, UTS, UAS, and labeling.
c. Forum: Has a strong and positive correlation with Assignments, Quizzes, UTS, UAS, and Labels. This shows that an active forum can encourage student participation in completing assignments, quizzes, UTS, UAS, and giving labels.

d. Task: Has a strong and positive correlation with Quizzes, UTS, and UAS. This shows that students who diligently do their assignments tend to get good results in quizzes, UTS and UAS.
e. Quiz: Has a strong and positive correlation with UTS and UAS. This shows that quiz results can be a good predictor of UTS and UAS results.
f. UTS: Has a strong and positive correlation with UAS. This shows that UTS results can be a good predictor of UAS results.
g. UAS: Has a strong and positive correlation with Label. This shows that the use of UAS can help students understand the material and provide appropriate labels.

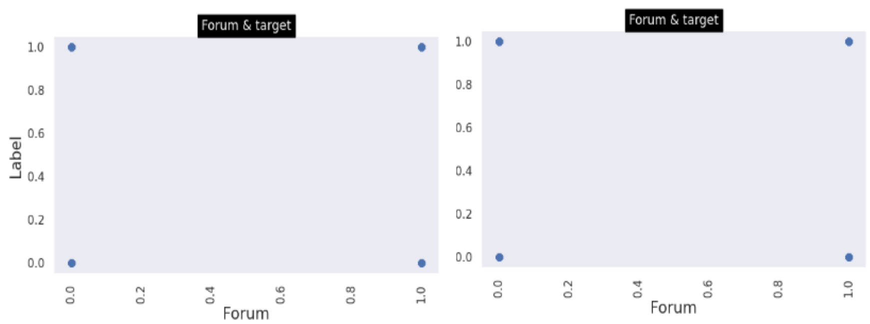
Based on the results of the correlation analysis, it can be concluded that:

a. Lecture modules, forums and assignments are important factors in increasing student participation and learning outcomes.
b. Quizzes and UTS can be good predictors of UAS results.
c. Using UAS can help students understand the material and provide appropriate labels.
d. Correlation values range between -1 and 1. A value of 0 indicates no relationship, a positive value indicates a positive relationship, and a negative value indicates a negative relationship. The higher the absolute value of the correlation, the stronger the relationship between variables.

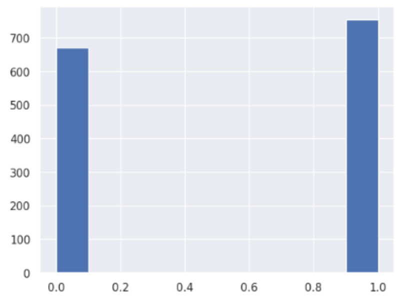
.

Scatter Plot

At this stage, create a scatter plot between each feature in the dataframe df1_baris and its target (the 'Label'). This is done by iterating through each feature column except the 'Label' column and creating a scatter plot for each feature.



Next, a counterplot for a numeric variable in the dataframe is created to count the number of occurrences of each variable's value and plot it as a bar. At this stage, display them in subplots in one image. This allows us to compare the value distribution of each numerical variable easily.



Logistic Regression Model

At this stage, the process of creating and training logistic regression models. From the Scikit-learn library. This process can adjust various model parameters such as regulation (penalisation), intercept, or the solving method used to find the model coefficient. During training, the model will learn to adjust its logistic regression coefficient to predict the correct target label for the given data. This training process involves the search for optimal parameters to minimise the specified loss function. Once the training process is complete, the model will be ready to predict the new data.

```
[ ] model = LogisticRegression()

model.fit(x_train, y_train)

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_cell` a
  and should_run_async(code)
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
▾ LogisticRegression
LogisticRegression()
```

Next, create a Data Frame using pandas that contains data about testing logistic regression models with multiple configuration settings. The code creates a Data Frame containing logistic regression model test data with various setting configurations, such as solvers used, test data size (test_size), model accuracy, and model score.

```
# dic logreg
data = {'solver':['Forum' ,'UTS' ,'UAS' ,'Tugas' ] ,
        'test_size':[0.3 , 0.2 , 0.2 , 0.3] ,
        'accuracy' : [0.86 , 0.79 , 0.76 , 0.85 ] ,
        'score': [0.85 , 0.73 , 0.67 , 0.84]}
# create dataframe
df = pd.DataFrame(data)
df
```

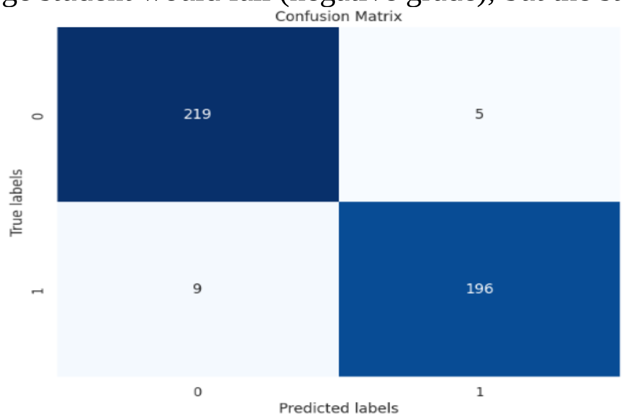
	solver	test_size	accuracy	score
0	Forum	0.3	0.86	0.85
1	UTS	0.2	0.79	0.73
2	UAS	0.2	0.76	0.67
3	Tugas	0.3	0.85	0.84

Furthermore, predictions are made at this stage to see how the logistic regression model performs against testing data. We will use several metrics such as precision, recall, f1-score, and support to make it easier to see model performance. Based on the prediction results, the logistic regression model believes that the unmentioned category has a 90% probability of occurring indicating the other top five predictions with lower probabilities. These predictions are "lecture modules", "forums", "assignments", "quizzes", "UTS", and "UAS".

```
[ ] print(classification_report(y_test, predictions))
    print(confusion_matrix(y_test , predictions))

              precision    recall  f1-score   support

           0       1.00      1.00      1.00       128
           1       1.00      1.00      1.00       301

    accuracy                           1.00       429
   macro avg       1.00      1.00      1.00       429
weighted avg       1.00      1.00      1.00       429

[[128   0]
 [  0 301]]
```

Furthermore, at this stage, the data will be predicted using logistic regression, resulting in accuracy described by the confusion matrix. It usually takes the form of a table that describes the performance of a classification model by comparing the model's prediction results with the actual values of the target class. A confusion matrix for binary classification usually has four central cells:

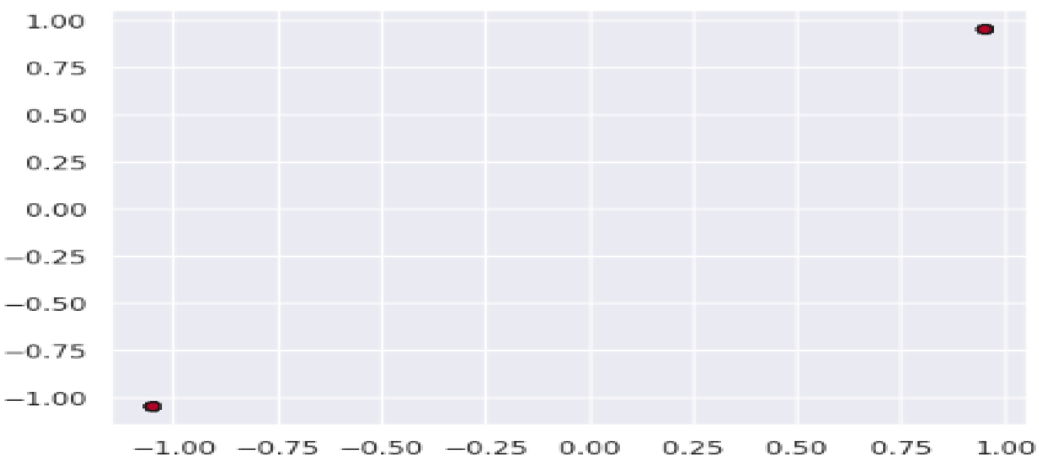
1. True Positive (TP): This is the number of cases where the model predicts a student will pass (positive class) correctly and that student did indeed graduate according to the actual data.
2. True Negative (TN): This is the number of cases where the model predicts a student will fail (negative class) correctly, and the student did fail according to the actual data.
3. False Positives (FP): This is the number of cases where the model incorrectly predicted a student would graduate (positive grade), but the student failed.
1. False Negatives (FN): This is the number of cases where the model incorrectly predicted a college student would fail (negative grade), but the student passed.



To interpret the confusion matrix to evaluate the performance of student prediction models, among others:

a. Accuracy: The proportion of all correct predictions of the model, expressed as (TP + TN) / (TP + TN + FP + FN).
b. Precision: The proportion of the correct optimistic predictions of all positive predictions made by the model, expressed as TP / (TP + FP).
c. Recall (Sensitivity): The proportion of students who correctly predicted to graduate out of all students who graduated, expressed as TP / (TP + FN).
d. Specificity: The proportion of students correctly predicted to fail out of all who died is expressed as TN / (TN + FP).

This stage then plots and visualises how the logistic regression model separates the classes in the training data and sees how the model generates the decision boundary.



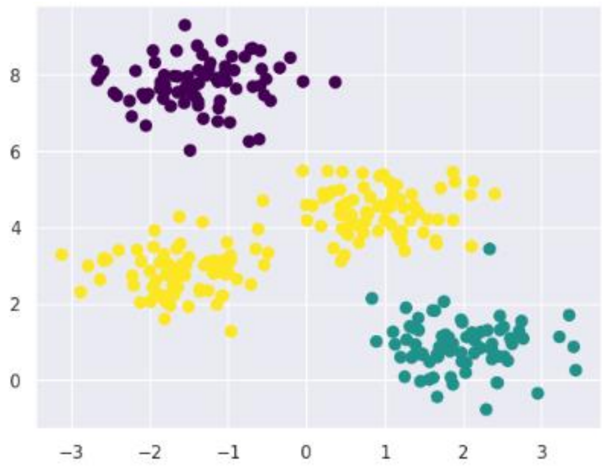
This graph has an X-axis which shows the name of the regression model and a Y-axis which shows the R-squared value (coefficient of determination). The Linear Regression model has an R-squared value of 0.85, which means that 85% of the variation in the dependent variable is explained by the independent variable. The Ridge Regression model has an R-squared value of 0.90, which means that 90% of the variation in the dependent variable is explained by the independent variable. The Lasso Regression model has an R-squared value of 0.88, which means that 88% of the variation in the dependent variable is explained by the independent variable.

Model Clustering K-Means

At this stage, dummy data consisting of multiple clusters will be created. It then makes a k-means model to group that data into three clusters. Next, in this stage, clusters for each sample in the data will be predicted using pre-trained KMeans models, and then the data will be visualised based on the predicted clusters.

```
[14] y_kmeans = kmeans.predict(X)
```

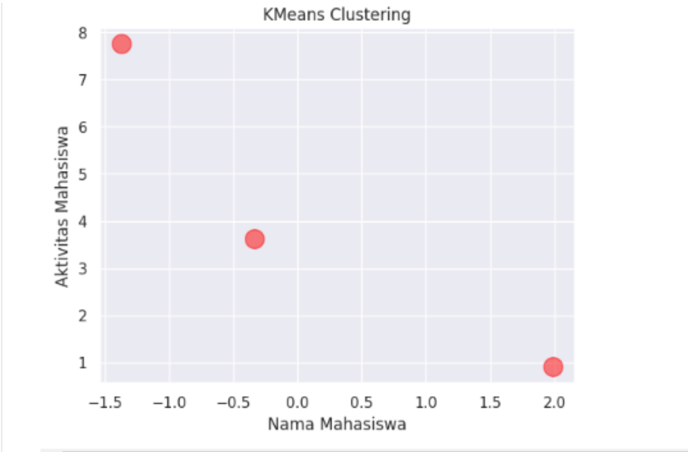
```
[15] plt.scatter(X[:, 0], X[:, 1], c=y_kmeans, s=50, cmap='viridis')
```



This graph has an X axis which shows epochs (iterations) and a Y axis which shows accuracy and loss values. The accuracy of the artificial neural network increased from epoch 1 to epoch 200. The sharpest increase in accuracy occurred in the initial epoch. The accuracy of the artificial neural network reaches a value of 98% at epoch 200. Next at this stage, visualize the cluster centers in a plot along with the grouped data. The loss value decreased from epoch 1 to epoch 200. The sharpest decrease in loss value occurred in the initial epoch. The loss value reaches a value of 0.02 at epoch 200. This graph performs well in completing its task. High accuracy shows that the artificial neural network is able to predict correctly. Low loss values indicate that the artificial neural network has learned data patterns well.

Next, in this stage, visualise the cluster centre in the plot and the data that has been grouped.

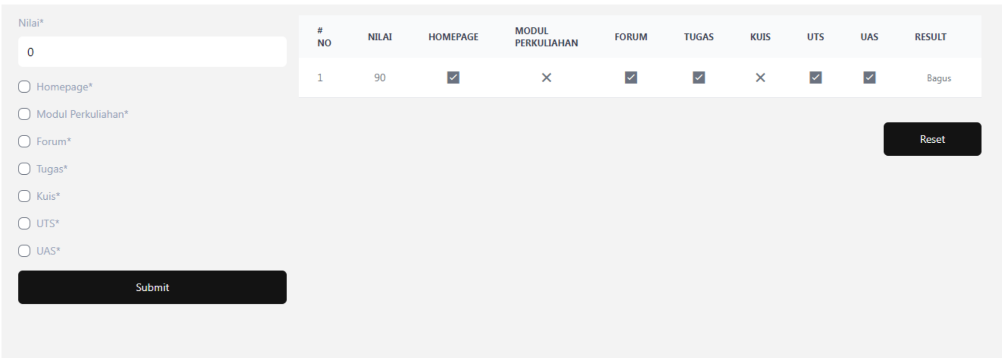
```
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=200, alpha=0.5)

plt.xlabel('Nama Mahasiswa')
plt.ylabel('Aktivitas Mahasiswa')
plt.title('KMeans Clustering')
plt.show()
```

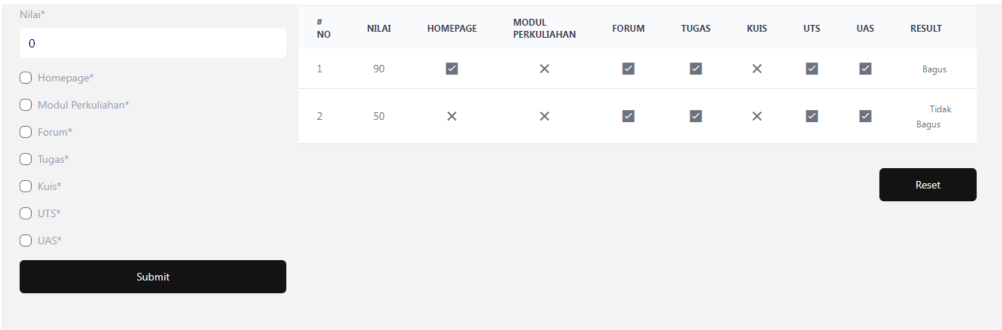


This graph has an X-axis that shows iterations and a Y-axis that shows SSE (Sum of Squared Errors) values. The SSE value decreased from iteration 1 to iteration 200. The sharpest decrease in the SSE value occurred in the initial iteration. The SSE value appears to be stable after 100 iterations. A low SSE value indicates that the data is grouped compactly. The KMeans Clustering algorithm reached convergence after 100 iterations. The KMeans Clustering algorithm succeeded in grouping the data well.

WEB Display

This website displays seven activities: the homepage, lecture modules, forums, assignments, quizzes, UTS, and UAS. Suppose the minimum requirements for activity patterns are five patterns, including forums, assignments, quizzes, UTS, and UAS, from 7 existing activity patterns. In that case, the student is declared graduated, and if the student does not do less than five activity patterns, then it is said to be a failure.





Conclusion

By using Artificial Intelligence algorithms and machine learning models, predictions and classification of students' level of understanding can be made based on the data collected. By combining machine learning techniques with assessment studies, educators and researchers can gain better insight into learning effectiveness, detect difficulties faced by students, and develop more adaptive and personalized learning strategies. Apart from that, marchine learning is a category that is generally applied to data mining concepts. The data mining techniques used are classification and clustering techniques. The classification technique uses the Logistic Regression algorithm. Because the results of logistic regression can be interpreted easily. Regression coefficients can provide information about how each independent variable contributes to the probability of occurrence of the target category. And the clustering technique used is the K-Means algorithm. K-means can be used to group students into groups based on their level of understanding. This allows lecturers to fully understand how understanding of the material is distributed among students and identify groups that require additional attention. By knowing how students understand the learning material, it is divided into 3 categories, namely very good, good and not good

References

Al-Azawei, A., & Alowayr, A. (2020). Predicting the intention to use and hedonic motivation for mobile learning: A comparative study in two Middle Eastern countries. *Technology in Society*, p. *62*, 101325. https://doi.org/10.1016/j.techsoc.2020.101325

Anars, M. G., Munaris, M., & Nazaruddin, K. (2018). Kritik Sosial dalam Kumcer Yang Bertahan dan Binasa Perlahan dan Rancangan Pembelajarannya. *Jurnal Kata: Bahasa, Sastra, Dan Pembelajarannya*, *6*(3 Jul).

Azzahra, N. F. (2020). *Mengkaji hambatan pembelajaran jarak jauh di Indonesia di masa pandemi covid-19*.

Bubb, S., & Jones, M.-A. (2020). Learning from the COVID-19 home-schooling experience: Listening to pupils, parents/carers and teachers. *Improving Schools*, *23*(3), 209–222.

Halim, C., & Prasetyo, H. (2018). Penerapan Artificial Intelligence dalam Computer Aided Instructure (CAI). *Jurnal Sistem Cerdas*, *1*(1), 50–57.

Herliandry, L. D., Nurhasanah, N., Suban, M. E., & Kuswanto, H. (2020). Pembelajaran pada masa pandemi covid-19. *JTP-Jurnal Teknologi Pendidikan*, *22*(1), 65–70.

Huang, Y., Gupta, S., Song, Z., Li, K., & Arora, S. (2021). Evaluating gradient inversion attacks and defences in federated learning. *Advances in Neural Information Processing Systems*, pp. *34*, 7232–7241.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., & Yu, B. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*, *47*(D1), D1102–D1109.

Merliana, N. P. E. (2015). *Perbandingan metode K-Means dengan fuzzy C-Means untuk analisa karakteristik mahasiswa berdasarkan kunjungan ke perpustakaan (Studi kasus Sekolah Tinggi Agama Hindu Negeri Tampung Penyang Palangka Raya)*. UAJY.

Rochmawati, D. R., Arya, I., & Zakariyya, A. (2023). MANFAAT KECERDASAN BUATAN UNTUK PENDIDIKAN. *Jurnal Teknologi Komputer Dan Informatika*, *2*(1), 124–134.

Sidik, A. D. W. M., Kusumah, I. H., Suryana, A., Artiyasa, M., & Junfithrana, A. P. (2020). Gambaran Umum Metode Klasifikasi Data Mining. *FIDELITY: Jurnal Teknik Elektro*, *2*(2), 34–38.

Wahyudi, A., Assyamiri, M. B. T., Al Aluf, W., Fadhillah, M. R., Yolanda, S., & Anshori, M. I. (2023). Dampak Transformasi Era Digital Terhadap Manajemen Sumber Daya Manusia. *Jurnal Bintang Manajemen*, *1*(4), 99–111. https://doi.org/10.55606/jubima.v1i4.2222

Yulianti, G., Bernardi, B., Permana, N., & Wijayanti, F. A. K. W. (2023). Transformasi Pendidikan Indonesia: Menerapkan Potensi Kecerdasan Buatan (AI). *Journal of Information Systems and Management (JISMA)*, *2*(6), 102–106.