DECODING DIGITAL LEARNING: A CART ANALYSIS INSIGHT

Sri Harini ¹, Angga Dwi Mulyanto²*, Fachrur Rozi³

¹Mathematic Study Program, State Islamic University of Maulana Malik Ibrahim Malang
Gajayana Street No 50, Malang, 65144, Indonesia
Corresponding author’s e-mail: *angga.dwi.m@mat.uin-malang.ac.id

ABSTRACT

In today’s digital era, online education has become an integral component of the global educational system. However, the quality of providers, mastery of online learning media, and the capability of teachers in delivering content significantly influence students’ comprehension. Understanding how these factors impact students is crucial for optimizing online educational experiences. This article examines the effect of provider quality, mastery of online learning media, and teacher’s capability in teaching on students’ understanding in an online learning environment using Classification and Regression Trees (CART). The study focused on the active students of UIN Maulana Malik Ibrahim Malang, totaling 18,104 based on data from PDDIKTI. Due to the vastness of this population, stratified random sampling was employed. The sample size was determined using the Kock and Hadaya method, resulting in a minimum sample of 271. Each sample in the stratum was taken based on the proportion of the population across 19 departments, with the Department of Psychology having the largest sample size of 20 students and the Department of Library and Information Science the smallest with 4 students. The findings revealed that a teacher’s capability has a more profound influence on students’ understanding than mastery of online learning media, while provider quality showed no significant impact. To enhance the effectiveness of online learning, it is recommended that educational institutions invest in teacher training focused on effective teaching methodologies and technological utilization, and also ensure the selection of online learning platforms that cater to student needs.

How to cite this article:
1. INTRODUCTION

The effect of provider quality, mastery of online learning media, and teacher's capability in teaching on students' understanding refers to how these factors influence a student's understanding of the material being taught in an online learning environment [1]. Provider quality refers to the overall level of expertise and professionalism of the provider of the online learning material, such as a university or educational institution [2]. A high-quality provider is likely to have well-designed and effective learning materials, which can lead to a better understanding of the material for the students [3]. Mastery of online learning media refers to the level of proficiency and skill of the teacher in using the technology and tools that are used to deliver the online learning material [4]. Teachers who have a high level of mastery of online learning media are likely to be able to deliver the material in a more engaging and effective way, which can lead to a better understanding of the material for the students [5]. Teacher's capability in teaching refers to the teacher's ability to effectively convey and explain the material being taught [6]. A teacher with a high level of capability in teaching is likely to be able to provide clear and concise explanations, use effective teaching methods, and provide appropriate feedback, which can lead to a better understanding of the material for the students [7]. Overall, CART analysis can be used to examine the relationship between these factors and students' understanding of the material [8], in order to identify which factors have the greatest impact on students' understanding and use this information to improve the quality of online learning materials, teacher training programs, and the overall effectiveness of online learning.

CART (Classification and Regression Trees) is a machine learning algorithm commonly used in data mining and statistical analysis [8, 10, 11]. It is suitable to be used in this case because CART can handle both categorical and continuous data and large datasets. CART can also identify important variables that have the most influence on the outcome variable, which is crucial in identifying which factors have the greatest impact on students' understanding in an online learning environment [12]. Furthermore, CART can be used to generate easy-to-interpret decision trees, which can help to identify patterns and relationships in the data [13]. By using CART analysis in this study, researchers can identify the most influential factors that affect students' understanding and use this information to improve the quality of online learning materials, teacher training programs, and the overall effectiveness of online learning.

While several studies have explored factors affecting online learning, few have combined the examination of provider quality, mastery of online learning media, and teacher's capability within the context of the UIN Maulana Malik Ibrahim Malang student population. Moreover, utilizing CART analysis to decipher the interplay between these factors, especially within this specific academic institution, presents a novel approach to understand the nuances influencing students' comprehension in online learning environments.

Online learning has become a popular and increasingly prevalent mode of education in recent years [14]. However, the effectiveness of online learning is still a topic of debate [15]. One of the key factors that can affect the effectiveness of online learning is the quality of the provider, the mastery of online learning media, and the teacher's capability in teaching [16]. The purpose of this study is to examine the effect of provider quality, mastery of online learning media, and teacher's capability in teaching on students' understanding [17]. This study will use CART analysis to analyze the data collected from students in an online learning environment. The findings of this study will provide insights into how these factors influence students' understanding and will help to improve the overall effectiveness of online learning.

2. RESEARCH METHODS

In this research, we used the Classification and Regression Trees (CART) analysis method to examine the relationship between provider quality, mastery of online learning media, and teacher's capability in teaching on students' understanding. The data for this study was collected from a sample of students in an online learning environment. The CART analysis was performed using the Minitab Software. The CART analysis allowed us to identify the most important factors that influence students' understanding.

CART is a machine learning algorithm that can be used for both classification and regression tasks. It works by recursively partitioning the data into smaller and smaller subsets until each subset is homogeneous. The response variable is then predicted for each subset using a simple rule, such as the majority class for classification or the mean value for regression.
CART analysis (Classification and Regression Trees) is a method used to build predictive models in statistics and machine learning. This method combines predictor variable selection and data partitioning into several parts based on variable selection rules and predictor value thresholds. Here are the general steps to perform CART analysis in the context of regression [8]:

1. **Variable Selection and Data Splitting:**
   - Select the best variable from the dataset to split the data.
   - For each variable, calculate the variance reduction and select the variable with the highest variance reduction.

2. **Node Creation:**
   - Create a node with the selected variable and divide the data into two groups based on the value of this variable.
   - The value at the node is the mean value of the Response Variable for all samples in that node.

3. **Recursion:**
   - For each group generated from the split, recursively repeat steps 1 and 2 to create sub-trees.

4. **Terminating the Recursion:**
   - If the maximum depth of the tree is reached or the target variance within a node is less than a certain threshold, terminate the recursion and make that node a leaf node.

**Mathematical Explanation:**

1. **Variance Reduction:**
   - Variance is a measure of how far the values in a dataset are from the mean of the dataset.
   - It is calculated as:
   \[
   \text{Variance}(S) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{1}
   \]
   where:
   - \(S\): Represents the dataset.
   - \(y_i\): The value of the Response Variable for the \(i\)-th sample.
   - \(\bar{y}\): The mean value of the Response Variable over the dataset.
   - \(n\): The number of samples in the dataset.
   - Variance reduction is calculated as:
   \[
   \text{Reduction} = \text{Variance}(S) - \left( \frac{|S_1|}{|S|} \text{Variance}(S_1) + \frac{|S_2|}{|S|} \text{Variance}(S_2) \right) \tag{2}
   \]
   where \(S\) is the initial dataset, and \(S_1\) and \(S_2\) are the datasets resulted from the split.

2. **Node Creation and Data Splitting:**
   - Data is split into two groups based on the value of the selected variable.
   - The value at the node is the mean value of the Response Variable for all samples in that node.
   \[
   \text{Node Value} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{3}
   \]

**Calculating R-squared (Coefficient of Determination):**

R-squared is a statistical measure that represents the proportion of the variance for a dependent variable that’s explained by an independent variable or variables in a regression model.

It is calculated as:
\[
R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{4}
\]
where:
- \(SS_{\text{res}}\) (residual sum of squares) is the sum of the squares of the difference between the observed actual outcomes and the predictions made by the model.
- \(SS_{\text{tot}}\) (total sum of squares) is the total variance in the dependent variable.

For each prediction, calculate \(SS_{\text{res}}\) and \(SS_{\text{tot}}\) as follows:
\[ SS_{\text{res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \]  
\[ SS_{\text{tot}} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \]  
Where \( \hat{y}_i \) is the predictions made by the model for the \( i \)th sample.

To obtain the "variable importance" in CART, you can use a metric called "mean decrease impurity" or "mean decrease in accuracy". This metric measures the importance of a predictor variable in influencing the quality of data separation by the decision tree. A higher "variable importance" score indicates a greater importance of the variable in prediction.

Here are the general steps to obtain the "variable importance" in CART:

1. Set Up:
   - Begin with a dataset with \( p \) predictors \( X_1, X_2, \ldots, X_p \) and a response variable \( Y \).
   - Fit a regression tree to this data.

2. Impurity Measure for Regression:
   - For regression, the impurity of a node is typically measured by the residual sum of squares.
   \[ SS_{\text{res}}(t) = \sum_{i\in t}(y_i - \bar{y}_t)^2 \]  
   where \( t \) is the set of training samples that fall into node \( t \) and \( \bar{y}_t \) is the mean response for the samples in node \( t \).

3. Calculation of the Importance of a Predictor:
   - For each internal node that splits on predictor \( X_j \), compute the reduction in \( SS_{\text{res}} \) due to that split. This is given by:
   \[ \Delta SS_{\text{res}} = SS_{\text{res}}(\text{parent}) - [SS_{\text{res}}(\text{left child}) + SS_{\text{res}}(\text{right child})] \]  
   - Accumulate these reductions in \( SS_{\text{res}} \) for each predictor over all the internal nodes where it's used for splitting. This will give a raw importance score for each predictor.

4. Normalization:
   - To make the variable importance scores comparable and to ensure they sum up to 1, normalize them:
   \[ \text{Importance}(X_j) = \frac{\text{Raw Importance}(X_j)}{\sum_{j=1}^{p}\text{Raw Importance}(X_j)} \]  
   Where \( p \) is the number of predictors. The raw importance of a predictor variable in a regression tree is calculated based on the total reduction in the Residual Sum of Squares that results every time that predictor is used to split nodes in the tree. The raw importance of predictor \( X_j \) is computed by summing up the reductions in \( SS_{\text{res}} \) for all internal nodes that split on \( X_j \):
   \[ \text{Raw Importance}(X_j) = \sum_{\text{all nodes splitting in } X_j} \Delta SS_{\text{res}} \]  

5. Ranking:
   - Rank the predictors based on their normalized importance scores.

6. Interpretation:
   - A higher importance score indicates that a predictor is more influential in predicting the response variable, given the data and the specific tree that was fit.

Variable importance in CART doesn't necessarily imply causation. Additionally, while the importance can give insights about influential variables in the model, it doesn't provide detailed information about the nature or magnitude of the relationship between predictors and the response.

It is important to note that the "variable importance" scores in CART regression are relative and depend on the chosen metric. The commonly used metrics are "mean decrease impurity" and "mean decrease in accuracy," but there may be variations depending on the specific implementation or analysis goals.

In the questionnaire, there are 3 indicators used to evaluate Provider Quality, they are Internet Speed, Signal Stability and Delay. While for Mastery of Online Learning Media, there are 2 indicators used, they are Mastery of Learning Media from the Teacher’s Perspective and Mastery of Learning Media from the Student’s Perspective. Meanwhile, for the variable of Teacher’s Capability, it is measured using 5 indicators,
they are Clarity of Material Presentation, Ability to Answer Questions, Ability to Make Students Active, Ability to Manage Time in Teaching, and Suitability of Learning Plan with Realization of Learning. Lastly, for the dependent variable Teaching on Students' Understanding is measured directly.

In this study, data was obtained from primary sources by distributing questionnaires to students. The population in this study is active students of UIN Maulana Malik Ibrahim Malang. Data on many populations obtained from PDDIKTI with a total of 18104. Due to the large number of students, we used sampling. Sample size was measured using the Kock and Hadaya method with a Power = 0.95, Alpha = 0.05, and Absolute Beta Minimal = 0.2, resulting in a minimum sample of 271. As for sampling, we used a stratified random sampling technique, using the 'Department' as the stratification variable. Each sample in the stratum was taken based on the proportion of the population, resulting in the sample needs, as can be seen in Table 1.

<table>
<thead>
<tr>
<th>Code</th>
<th>Department</th>
<th>Number of Students</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>62201</td>
<td>Accounting</td>
<td>615</td>
<td>10</td>
</tr>
<tr>
<td>79203</td>
<td>Arabic Language and Literature</td>
<td>977</td>
<td>15</td>
</tr>
<tr>
<td>46201</td>
<td>Biology</td>
<td>738</td>
<td>12</td>
</tr>
<tr>
<td>48201</td>
<td>Pharmacy</td>
<td>560</td>
<td>9</td>
</tr>
<tr>
<td>45201</td>
<td>Physics</td>
<td>395</td>
<td>6</td>
</tr>
<tr>
<td>74234</td>
<td>Sharia Economic Law (Mu'amalah)</td>
<td>927</td>
<td>14</td>
</tr>
<tr>
<td>74230</td>
<td>Islamic Family Law (Ahwal Syakhshiyah)</td>
<td>1007</td>
<td>16</td>
</tr>
<tr>
<td>74235</td>
<td>Constitutional Law (Siyasah)</td>
<td>544</td>
<td>9</td>
</tr>
<tr>
<td>76231</td>
<td>Quranic Studies and Exegesis</td>
<td>351</td>
<td>6</td>
</tr>
<tr>
<td>47201</td>
<td>Chemistry</td>
<td>624</td>
<td>10</td>
</tr>
<tr>
<td>61201</td>
<td>Management</td>
<td>1245</td>
<td>19</td>
</tr>
<tr>
<td>86231</td>
<td>Islamic Education Management</td>
<td>473</td>
<td>8</td>
</tr>
<tr>
<td>44201</td>
<td>Mathematics</td>
<td>597</td>
<td>9</td>
</tr>
<tr>
<td>86208</td>
<td>Islamic Religious Education</td>
<td>988</td>
<td>15</td>
</tr>
<tr>
<td>88204</td>
<td>Arabic Language Education</td>
<td>696</td>
<td>11</td>
</tr>
<tr>
<td>11201</td>
<td>Medical Education</td>
<td>218</td>
<td>4</td>
</tr>
<tr>
<td>86232</td>
<td>Early Madrasah Teachers Education</td>
<td>697</td>
<td>11</td>
</tr>
<tr>
<td>87220</td>
<td>Social Science Education</td>
<td>678</td>
<td>11</td>
</tr>
<tr>
<td>86233</td>
<td>Early Childhood Islamic Education</td>
<td>293</td>
<td>5</td>
</tr>
<tr>
<td>61206</td>
<td>Sharia Banking</td>
<td>617</td>
<td>10</td>
</tr>
<tr>
<td>71202</td>
<td>Library and Information Science</td>
<td>247</td>
<td>4</td>
</tr>
<tr>
<td>73201</td>
<td>Psychology</td>
<td>1317</td>
<td>20</td>
</tr>
<tr>
<td>79202</td>
<td>English Literature</td>
<td>1184</td>
<td>18</td>
</tr>
<tr>
<td>88203</td>
<td>English Language Teaching</td>
<td>300</td>
<td>5</td>
</tr>
<tr>
<td>84202</td>
<td>Mathematics Teaching</td>
<td>268</td>
<td>5</td>
</tr>
<tr>
<td>23201</td>
<td>Architectural Engineering</td>
<td>698</td>
<td>11</td>
</tr>
<tr>
<td>55201</td>
<td>Informatics Engineering</td>
<td>850</td>
<td>13</td>
</tr>
</tbody>
</table>

Based on Table 1, it can be seen that there are 19 departments at UIN Maulana Malik Ibrahim Malang. The Department of Psychology has the largest sample size of 20 students, while the smallest sample size is in the Department of Library and Information Science with 4 students.

3. RESULTS AND DISCUSSION

The selection of optimal nodes is done by using R-squared. When the R-squared value starts to become constant, that is where the optimal number of nodes is. Based on Figure 1, it can be seen that the
optimal number of terminal nodes is 5 with an R-squared value of 34.61%. In the R-squared versus the number of terminal nodes plot, it is evident that the value of R-squared is higher at node 9 than at node 5. However, the “Optimal” point is marked at node 5 with a value of 34.61. The decision for such an “optimal” point is usually influenced by considerations beyond just the R-squared value. Firstly, higher R-squared values, as seen at node 9, might lead to overfitting, where the model becomes too complex and performs poorly on unseen data. Node 5 may provide a good balance between model complexity and its generalization capability. Secondly, in decision tree pruning techniques, “cost-complexity pruning” can determine a balance between model error and its complexity, and node 5 might offer a favorable trade-off in this context. Cross-validation is another potential factor: even if node 9 has a higher R-squared on one data subset, node 5 might deliver more consistent performance across different data subsets. The R-squared is indeed on the lower side. This might be due to missing important variables in this study, like environmental, social, student psychology, and perhaps other factors not mentioned yet. This will be an interesting note for future research development.

![R-squared vs Number of Terminal Nodes Plot](image)

**Figure 1. The Selection of Optimal Terminal Nodes**

Based on the Tree Diagram in Figure 2, there are 5 Terminal Nodes with the following explanations:

1. Terminal Node 1 shows a group with Teachers' Capability values less than or equal to 21.5. The mean value for Teaching on Students' Understanding is 1, which is still below the overall mean value (6.70280).
2. Terminal Node 2 shows a group with Teachers' Capability values between 21.5 and 32.5. The mean value for Teaching on Students' Understanding is 5.3, which is still below the overall mean value (6.70280) but better than Terminal Node 1.
3. Terminal Node 3 shows a group with Teachers' Capability values between 32.5 and 43.5 and Mastery of Online Learning Media values less than or equal to 13.5. The mean value for Teaching on Students' Understanding is 5.8, which is still below the overall mean value (6.70280) but approaches the overall mean value.
4. Terminal Node 4 shows a group with Teachers' Capability values between 21.5 and 32.5 and Mastery of Online Learning Media values greater than 13.5. The mean value for Teaching on Students' Understanding is 7.08, which is above the overall mean value (6.70280).
5. Terminal Node 5 shows a group with Teachers' Capability values above 43.5. The mean value for Teaching on Students' Understanding is 8.30, which is above the overall mean value (6.70280) and above the mean value of Terminal Node 4.
The relative importance graph in Figure 3 shows that the most important predictor variable is Teacher’s Capability. If predictor variable contributions are sorted from the top, Teacher’s Capability is 100%, then we can compare other variables to Teacher’s Capability to determine their level of importance. Thus, we can focus on the most important predictors. The following list explains the next most important variables in this analysis:

1. Mastery of Online Learning Media has an importance of about 54.3% compared to Teacher’s Capability.

2. Provider Quality has an importance of about 14% compared to Teacher’s Capability.

Although these results include three variables with positive importance, their relative ranking provides information on how many variables need to be controlled or monitored in this case. A drastic decrease in relative importance value from one variable to the next can help make decisions about which variables to control or monitor. For example, in this case, the three variables have relatively distant importance values from one another. From Teacher’s Capability to Mastery of Online Learning Media, there is a decrease of 45.6%. While from Mastery of Online Learning Media to Provider Quality, there is a decrease of 40.3%. This shows that the most important factor influencing Teaching on Students’ Understanding is Teacher’s Capability. The other two variables still have much lower importance in influencing Teaching on Students’ Understanding compared to Teacher’s Capability.
Based on the results of the CART analysis above, it can be seen that increasing the value of Teacher's Capability will increase the value of Teaching on Students' Understanding. This is in line with the findings of Noviana and Solichin [18]; Mazzanti and Karsli-Calamak [19]; Innayah [20]; Lindell [21]; Kokkinos and Gakis [22]; Orakcı [23]; and Jager, Denessen, Cillessen and Meijer [24], where it was found that Teacher's Capability has a positive effect on Teaching on Students' Understanding. The same goes for Mastery of Online Learning Media, as the value of Mastery of Online Learning Media increases, it will also increase the value of Teaching on Students' Understanding. This is in line with the findings of Innayah [20]; Wei and Chou [25]; Nuraisyah, Harahap, and Harahap [26]; and Mustofa [27], where it was found that Mastery of Online Learning Media has a positive effect on Teaching on Students' Understanding.

Mastery of Online Learning Media is one of the factors that can affect Teaching on Students' Understanding. This study found that instructors who have good skills in using online learning media have better teaching performance and effectiveness in helping students understand the learning concept. This is because online learning media can provide various ways and variations of learning that meet the diverse needs of students. By mastering online learning media, instructors can deliver learning material more interactively and attractively, thereby increasing students' interest in learning and understanding.

In addition, Teacher's Capability also has an effect on Teaching on Students' Understanding. Instructors who have good and effective teaching skills can deliver learning material in a way that is easily understood by students. This study found that teachers who are able to create a positive learning environment, can motivate students to learn well, and overcome students' learning difficulties, tend to have better and more effective teaching performance in improving students' understanding. Therefore, there is a need to pay attention to improving teachers' teaching skills, such as skills in designing and delivering learning material, utilizing existing learning media, as well as ways to overcome various problems in the learning process. Teacher's capability plays a crucial role in influencing students' understanding compared to mastery of online learning media or provider quality. There are several reasons why this is the case.

Firstly, teachers are the facilitators and guides in the learning process. Their knowledge, expertise, and teaching skills greatly impact students' understanding. A capable teacher possesses a deep understanding of the subject matter, which allows them to explain complex concepts in a clear and engaging manner. They have the ability to adapt their teaching methods to cater to students' individual needs, ensuring that every student comprehends the material effectively.

Secondly, teachers provide direct interaction and personalized support to students. In a traditional classroom setting, teachers can gauge students' understanding through real-time feedback and adjust their teaching accordingly. They can identify areas where students may be struggling and offer additional explanations or examples to enhance comprehension. Such personalized attention is challenging to replicate solely through online learning media or provider platforms.

Thirdly, teachers play a vital role in motivating and inspiring students. Their enthusiasm, passion, and dedication create a positive learning environment that fosters student engagement and active participation. Teachers can establish meaningful connections with students, encouraging them to strive for excellence and develop a genuine interest in the subject matter. These motivational aspects are often difficult to replicate in an online learning environment or through generic provider platforms.
Furthermore, teachers possess pedagogical expertise and instructional strategies that enhance learning outcomes. They are skilled in employing various teaching techniques, such as interactive discussions, hands-on activities, and formative assessments, to promote deep understanding and critical thinking. Through their experience, teachers can adapt the curriculum to meet the specific needs of their students and employ effective strategies to address any learning gaps.

Lastly, teachers serve as role models and mentors for students. They provide guidance beyond academic knowledge, nurturing students’ personal and social development. Teachers can instill values, foster good study habits, and encourage self-discipline, all of which contribute to a conducive learning environment and improved understanding.

While mastery of online learning media and provider quality are important factors in facilitating effective learning, the impact of a capable teacher cannot be understated. Their knowledge, teaching skills, personalized support, motivation, instructional strategies, and mentorship significantly contribute to students’ understanding and overall educational experience.

In light of the aforementioned findings, the novelty of this research lies in highlighting the paramount significance of Teacher's Capability in shaping students' understanding, even in the digital age where online learning tools are proliferating. While many modern studies have veered towards the efficacy of digital platforms and their intricate features, our research pivots the focus back to the human element - the teacher. The results underscore the timeless truth that while technological aids can supplement the learning process, it is the teacher's expertise, adaptability, and personal connection with students that remain the cornerstone of effective education. This revelation prompts educational institutions and policymakers to rethink and re-emphasize the value of investing in teacher development, even as the digital transformation in education continues.

4. CONCLUSIONS

It can be concluded that the factors that influence Teaching on Students' Understanding in this study are Mastery of Online Learning Media and Teacher's Capability. Teacher's Capability has a greater importance in influencing Teaching on Students' Understanding than Mastery of Online Learning Media. Meanwhile, Provider Quality does not affect Teaching on Students' Understanding.

One suggestion to improve the effectiveness of online learning based on the findings of this study is to focus on improving teacher's capability in teaching. This can be achieved through providing adequate training and resources to help teachers develop their skills in delivering online instruction and managing virtual classrooms. Additionally, it may be beneficial to provide ongoing support to teachers to help them adapt and improve their teaching methods as technology and online learning environments continue to evolve. Finally, further research can be conducted to explore other potential factors that may affect the effectiveness of online learning and identify additional areas for improvement.

ACKNOWLEDGMENT

Thank you to Maulana Malik Ibrahim State Islamic University of Malang for funding this research under the BOPTN 2022 research scheme.

REFERENCES


