Development of Teaching Performance Evaluation Application for Lecturers Using K-Nearest Neighbor Method with Manhattan Distance Approach

Annisa Helmina¹(⁎), Dedy Irfan², Fahmi Rizal³, Kasmita⁴
¹,²,³,⁴Universitas Negeri Padang, Padang, and Indonesia

Abstract
Based on the initial observation, there are several underlying issues that form the basis of this research. The teaching performance evaluation at Padang State University has limitations regarding the courses to be evaluated. Each student can only evaluate 5 (five) courses per semester, where these five courses are randomly selected by the system, allowing each student to evaluate different courses even in the same class. The evaluation of teaching performance at Padang State University is not specific to individual lecturers but to the courses. One course can be taught by several lecturers, so students evaluating the learning cannot provide assessments for each lecturer. This results in each lecturer not having their own performance results. Furthermore, the teaching performance evaluation at Padang State University does not have a classification for the filled evaluations, thus requiring a long time to calculate the final results. This study uses the Research and Development (R&D) method with the 4-D development model consisting of four stages: definition, design, development, and dissemination. The type of data used is primary data obtained from 3 media validators, the evaluation administration of Padang State University, and 46 students. The data analysis technique used is descriptive data analysis to describe the validity and practicality of the developed lecturer performance evaluation application. The results of this development study produced a lecturer performance evaluation with 202 training data, 47 test data, resulting in an accurate system with a precision value of 88.76%, a recall value of 89.93%, and a program accuracy value of 94.04%. The validity results of the web-based learning evaluation conducted by media experts obtained a score of 0.864 with a valid category. The practicality value of using web-based lecturer performance evaluation by students obtained a score of 87.42 with a practical category. Meanwhile, the practicality value obtained by the evaluation administration of Padang State University is 89.33 with a practical category.

Keywords: Evaluation, Assessment of Lecturer Performance, K-Nearest Neighbor, Manhattan Distance

(*) Corresponding Author:


INTRODUCTION

Information technology has penetrated various aspects of the education sector, including the preparation of learning tools, the implementation of teaching and learning processes, and the assessment of learning outcomes (Rudini & Saputra, 2022). This necessitates educators to leverage technology to address the shortcomings that may arise in conventional evaluation methods, particularly in
the final stage of learning evaluation (Nurjanah & Nurfitriani, 2020). In addition to the advancement of information technology, there has also been rapid progress in the development of applications (Ari Saktiono, 2019). In the field of education, the positive impact of the advancement of information technology is becoming increasingly evident, as its development has led to significant changes in the education sector (Asmawi et al., 2019). The education sector plays a crucial and strategic role as it serves as a means to enhance the quality of individuals (Eliyarti et al., 2020). A similar situation occurs in the context of higher education, where students as learners are considered mature individuals capable of rational and independent thinking (Eliyarti & Rahayu, 2019).

Higher education is the educational level pursued after secondary education (Rahmat Hidayat, 2022). In its role, higher education holds significant importance in producing individuals with high intellectual capabilities who ultimately contribute to the advancement of the nation (Taqwa et al., 2020). This is primarily due to the role of a teacher, who is an essential element in the context of higher education or the teaching-learning process (Multazam et al., 2018). Every teacher (lecturer) and university has the freedom to choose a teaching approach that suits the situation and subject matter encountered, both in theory and practice (Mexda & Mukhaiyar, 2021). To measure the achievement of learning outcomes, the use of appropriate assessment tools is necessary, such as evaluation methods that can be applied by lecturers, such as written tests that include various types of questions such as multiple choice, essay, true/false, matching, fill-in-the-blank, and similar types (Aripin, 2019).

The assessment process for a course can be conducted by conducting surveys through questionnaires given to students (Dewi et al., 2022). This questionnaire consists of several questions presented in multiple-choice format and provides students with the opportunity to provide comments (Lapi et al., 2021). The purpose of filling out the questionnaire is to obtain feedback on the evaluation of the performance or ability of the lecturer in teaching a particular course (Ramadhani & Fajarianto, 2020). The evaluation of lecturers' performance is conducted to optimize the teaching and learning activities in universities (Jaya et al., 2019). Evaluation can be defined as a structured, continuous, and comprehensive process aimed at controlling, ensuring, and determining the quality of various elements of learning (Hasanudin et al., 2021).

At Universitas Negeri Padang, learning evaluation is still limited to 5 courses, which means students do not have the freedom to evaluate all the courses they take. Courses at Universitas Negeri Padang may be taught by 2 to 3 lecturers, so the learning evaluation is not specific to one lecturer, making it difficult for students to provide accurate ratings as they may not know which lecturer is being evaluated. As a result, lecturers do not receive evaluations or feedback on their performance throughout the semester. The teaching performance evaluation at Universitas Negeri Padang also means that each student can only assess five randomly selected courses taken by the system, leading to potential discrepancies in the evaluations among students in the same class. This situation results in the suboptimal evaluation process of courses at Universitas Negeri Padang.
METHODS

The research method applied is research and development. The Research and Development (R&D) method is a research approach used to create specific products and test their performance (Okpatrioka, 2023). The development model utilizes the 4-D approach, consisting of four stages: Define, Design, Develop, and Disseminate (Nurmitasari et al., 2020). This model approach was chosen for this research because its process is structured and aligns with the underlying research issues (Rudini & Saputra, 2022). This research was conducted at the Informatics Study Program, Faculty of Engineering, Universitas Negeri Padang. The subjects of the study were undergraduate students majoring in Informatics at Universitas Negeri Padang who were attending classes in the January-June 2023 semester. Validation was carried out by three media expert lecturers. The media practicality test questionnaire was filled out by students and the learning evaluation organizers at Universitas Negeri Padang. The selection of the K-NN algorithm was made because K-NN is an efficient method for processing data on a large scale, capable of handling noisy training data, and demonstrates optimal performance (Putry & Sari, 2022). Manhattan distance is a calculation technique within the framework of distance space by applying the principle of absolute differences (Azis et al., 2021).

RESULTS & DISCUSSION

Define Stage

In order to obtain an overview of lecturer performance evaluation, the definition stage is conducted (Rajagukguk et al., 2021). During this stage, various analyses are conducted, including needs analysis, student analysis, concept analysis, and formulation of conventional objectives, to understand the development of lecturer performance evaluation in learning using the web-based K-Nearest Neighbor (K-NN) Manhattan Distance method.

The needs analysis is conducted to observe initial issues, identify barriers, problems, and on-field situations related to learning evaluation. The ultimate goal of needs analysis is to create solutions that fit the learning evaluation conditions. The student analysis is conducted to understand the characteristics of students, including their backgrounds, abilities, interests, motivations, and skills. Based on their intellectual development, students fall into the category of individuals with skills in utilizing digital technology and smartphones. Thus, the developed evaluation utilizes technology in the form of a website. Concept analysis is performed by identifying concepts that need to be considered in evaluating lecturer performance in learning, organized in a detailed and systematic manner. Furthermore, analysis is conducted to determine conventional objectives, where changes expected after using the developed lecturer performance evaluation in learning will be formulated.

Design Stage

This stage is often referred to as the design process (blueprint) (Gaja & Hendrik, 2023). In the analogy of building construction, before construction begins, there needs to be a design first that is put on paper. Therefore, the elements in the 4-D learning system are often represented in a diagram illustrating the stages from start to finish. The
programming method used is the K-Nearest Neighbor (K-NN) data mining algorithm with the designed system scheme. The workflow of the K-NN method and the system scheme used are as follows:

![Figure 1. Workflow of the K-Nearest Neighbor Method](image)

Skema sistem determining the evaluation results of lecturer performance in learning based on questionnaire data using the K-NN Manhattan Distance method is illustrated in the following Figure 2:

![Figure 2. Schematic of Lecturer Performance Evaluation System in Learning](image)

a) Initial Website Design
The initial design is a mandatory planning for web-based learning evaluation before it is validated and tested. The developed program characteristics include text questions, calculations, and evaluation of lecturer performance in web-based learning. The initial work stage involves creating a use case diagram, as follows:
Figure 3. Use Case Diagram

In this diagram, there are two actors (actors): admin and user. The admin can access the system's workflow by filling out questionnaire data based on predetermined choices, then view the evaluation results of lecturer performance in learning based on the filled data. The deep meaning of the user in this developed system is that students can only fill out evaluations of lecturer performance in learning.

b) Website Development

After the initial program design is implemented, it proceeds to the second stage, which is program development. The program development includes several menus such as questionnaire form pages, login, account data, course registration, training data, K-NN evaluation, calculation accuracy, and admin instructions regarding program usage. For the program details, please refer to the image below:

Figure 4. Form for Evaluating Lecturer Teaching Performance
The main page (Home-Questionnaire Form) will appear first when entering the system. On this main page, there is a form for evaluating lecturer performance in learning where users can enter data to be processed by the system later. Additionally, this main page also contains several other menus: "About the Application" and "Help" menus.

![Figure 5. Web-Based Learning Evaluation Login Page](image)

The "Login" page is the page used by the admin to log in to the system. If the email and password entered match the data in the database, the system will display the Dashboard. If they do not match, the system will prompt the user to re-enter the email and password.

c) Perhitungan Menggunakan K-NN

The K-Nearest Neighbor (K-NN) method is an algorithm used to classify objects based on training data by grouping new data based on their proximity to a certain number of nearest neighbors that are closest to the new data (Hasym & Susilawati, 2021). The K-Nearest Neighbor method is frequently applied in the classification process (Pranajaya & Febriansyah, 2021). The objective of this method is to categorize new objects based on their attributes and the training samples (Putri et al., 2021).

$$K - NN = \left( \frac{S_1 \times W_1}{W_1} + (S_2 + W_2) + \cdots (S_n + W_n) }{(W_1 + W_2 + \cdots W_n)} \right)$$

From the K-Nearest Neighbor equation mentioned above, there are two variables, namely S representing similarity (level of similarity) with a value of 1 indicating similarity and a value of 0 indicating difference. Meanwhile, W is the weight or coefficient on each symptom. The K-Nearest Neighbor method is applied with the intention of increasing the level of similarity values by applying indexing between Sorgenfrei and K-NN, thus resulting in a more optimal level of similarity for the investigated objects (Ramadhani & Fajariant, 2020).

![Figure 6. Example of K-NN calculation for one of the lecturers](image)
After that, manual search is conducted, or item-by-item search based on the entered data is then compared with the existing training data, so the approach taken by Manhattan Distance will facilitate the search. Manhattan Distance search, which sorts the data closest to the existing training data. For more details, please refer to Figure 7 below:

![Figure 7](image1)

**Figure 7. Ordered Distance After Search Guidance**

After the data is sorted based on the similarity to the training data, the 3 nearest distances will be selected. In the example above, for the 3 nearest data points or K=3, it can be seen in Figure 8 below:

![Figure 8](image2)

**Figure 8. Closest Distances (K=3)**

After obtaining the 3 closest distances with the help of the Manhattan Distance approach, the next step is to find the similarity or similarity by comparing the three closest distances obtained. For more details, please refer to Figure 9 below:

![Figure 9](image3)

**Figure 9. Similarity of K=3**
So, after obtaining the similarity from the entered data, the classification results will be obtained. For the example above, the classification results can be seen in Figure 10 below:

![Figure 10. Classification Results](image)

**a) Measurement Accuracy**

The accuracy evaluation method involves the use of a confusion matrix. The confusion matrix is a table that shows the classification between the number of correct test data and the number of incorrect test data (Normawati & Prayogi, 2021). In the development of learning evaluation using the K-Nearest Neighbor (K-NN) Manhattan Distance approach, there are 202 training data and 47 test data. So, when the test data is manually calculated using the K-NN method, the resulting data is as follows:

![Figure 11. Confusion Matrix](image)

Confusion Matrix if written in a calculation formula for example below:

1. **Precision**

   Precision is used to illustrate the requested rate between the input data and the prediction results generated by the model (Fauziningrum & Suryaningsih, 2021).

   \[
   \text{Precision} = \frac{TP}{FP + TP} \times 100\% 
   \]

   \[
   \text{very good} = \frac{5}{2+5} \times 100\% = 71.42\% 
   \]

   \[
   \text{good} = \frac{26}{4+26} \times 100\% = 86.67\% 
   \]

   \[
   \text{pretty good} = \frac{6}{14+6} \times 100\% = 85.71\% 
   \]

   \[
   \text{not good} = \frac{2}{0+2} \times 100\% = 100\% 
   \]

   \[
   \text{very not good} = \frac{1}{0+1} \times 100\% = 100\% 
   \]

   **Average Precision** = \[
   \frac{71.42+86.67+85.71+100+100}{5} \times 100\% = 88.76\% 
   \]
2. Recall

Recall or sensitivity indicates the model's ability to identify information correctly (Fauzineringrum & Suryaningsih, 2021).

\[
Recall = \frac{TP}{FN + TP} \times 100\%
\]

very good = \frac{5}{0+5} \times 100\% = 100\%
good = \frac{26}{3+26} \times 100\% = 89,65\%
Pretty good = \frac{6}{4+6} \times 100\% = 60\%
not good = \frac{2}{0+2} \times 100\% = 100\%
very not good = \frac{1}{0+1} \times 100\% = 100\%

Average Precision = \frac{\text{very good} + \text{good} + \text{Pretty good} + \text{not good} + \text{very not good}}{5} \times 100\% = 89,93\%

3. Accuracy

Accuracy is beneficial in understanding how far the classification prediction accuracy corresponds to the actual classes (Adek et al., 2020).

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%
\]

very good = \frac{5+40}{5+40+2+0} \times 100\% = 95,74\%
good = \frac{26+14}{26+14+4+3} \times 100\% = 85,11\%
Pretty good = \frac{6+36}{6+36+1+4} \times 100\% = 89,36\%
not good = \frac{2+45+0+0}{2+45+0+0} \times 100\% = 100\%
very not good = \frac{1+46}{1+46+0+0} \times 100\% = 100\%

\[
Rata - Rata Accuracy = \frac{95,74+85,11+89,36+100+100}{5} \times 100\% = 94,04\%
\]

Develop Stage

The development stage is where the plan or design that has been created is implemented into reality. In this stage, all components needed or supporting the creation of the application must be carefully prepared. This research aims to design learning evaluation, so the final product must go through a structured development process from the initial stage to the final development stage. The main goal in this stage is to produce a developed product that has been adjusted based on input from validators, resulting in a valid web-based learning evaluation ready for testing.

a. Validity Testing

The results of the assessment for each aspect provided by the validator are analyzed using Aiken’s V formula, which is outlined as follows:

\[
V = \sum s / [n (c - 1)]
\]

Information:

V = Validity coefficient
s = r – lo
lo = The lowest validity value
c = The highest validity value
r = The value provided by a validator
Table 1. Response Criteria Analysis

<table>
<thead>
<tr>
<th>No</th>
<th>Achievement Level</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥ 0,667 – 1,00</td>
<td>Valid</td>
</tr>
<tr>
<td>2</td>
<td>≤ 0,667</td>
<td>Not valid</td>
</tr>
</tbody>
</table>

The results obtained are validation values against the design of the produced product. The validation results based on the validation questionnaire of the developed web-based learning evaluation can be seen in the following table:

Table 2. Expert Validation Results

<table>
<thead>
<tr>
<th>No</th>
<th>Aspect</th>
<th>V</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ease of Use and Navigation</td>
<td>0,806</td>
<td>Valid</td>
</tr>
<tr>
<td>2</td>
<td>Visual Appearance</td>
<td>0,833</td>
<td>Valid</td>
</tr>
<tr>
<td>3</td>
<td>Media Integration</td>
<td>0,958</td>
<td>Valid</td>
</tr>
<tr>
<td>4</td>
<td>Media Benefits</td>
<td>0,833</td>
<td>Valid</td>
</tr>
<tr>
<td>5</td>
<td>Technical Quality</td>
<td>0,889</td>
<td>Valid</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>0,864</td>
<td>Valid</td>
</tr>
</tbody>
</table>

Based on the table above, the validity test results for media obtained a validation value of 0.864 > 0.667, so the web-based learning evaluation is categorized as valid. Although based on statistical tests, the validation by media experts is already valid, there are still some minor revisions needed.

b. Practicality Testing

The practicality analysis of web-based learning evaluation based on user data is depicted using the frequency data analysis technique obtained from observation sheets on the questionnaire. The practicality analysis of web-based learning evaluation questionnaire uses the following formula:

\[
\text{Percentage} = \frac{\text{Sum of responden's score}}{\text{Sum of the maximum score}} \times 100\%
\]

The percentage achievement level of respondent practicality categories is presented in Table 3 below:

Table 3. Percentage of Practicality Level

<table>
<thead>
<tr>
<th>No</th>
<th>Achievement Level(%)</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90 – 100</td>
<td>Very Practical</td>
</tr>
<tr>
<td>2</td>
<td>80 – 89</td>
<td>Practical</td>
</tr>
<tr>
<td>3</td>
<td>65 – 79</td>
<td>Fairly Practical</td>
</tr>
<tr>
<td>4</td>
<td>55 – 64</td>
<td>Not Quite Practical</td>
</tr>
<tr>
<td>5</td>
<td>0 – 54</td>
<td>Impractical</td>
</tr>
</tbody>
</table>

The practicality test was conducted by 46 students from 2 classes, namely INF1 class with 22 students and INF2 class with 24 students, in the Informatics Engineering Program of the Faculty of Engineering, Universitas Negeri Padang, who were taking courses in the semester from January to June 2023. The initial testing phase involved distributing the web-based learning evaluation link to students via the class WhatsApp groups with the link https://evaluasipb.freetechcreative.com/. Then, students were asked to fill out the questionnaire on their respective devices. After opening the web-based learning evaluation website, students were then asked to fill out a practicality questionnaire to assess the practicality of using the web-based learning evaluation website that had been developed. The results of students' responses to the practicality of using the web-based learning evaluation for assessing the performance of lecturers obtained a
practicality score of 87.42, categorized as practical. The results of the administrator's responses to the use of the web-based learning evaluation for assessing the performance of lecturers obtained a practicality score of 89.33, also categorized as practical.

Disseminate Stage
After designing the web-based learning evaluation and obtaining valid and practical results, it can be deemed ready for dissemination. The web-based learning evaluation, once prepared, can be distributed or utilized by students at Universitas Negeri Padang. This dissemination phase is carried out with the aim of assessing the implementation of using the web-based learning evaluation by both administrators and students, thus making it a tool to enhance the quality assurance of Universitas Negeri Padang.

CONCLUSION
This research employs a quantitative research method using the R&D development type and follows the 4-D development model. With 202 training data and 47 test data, the study results in a web-based learning evaluation achieving an accuracy of 94.04%, with precision at 88.76% and recall at 89.93%. The validity of the learning performance evaluation conducted by three media experts obtains a score of 0.864, categorized as valid. The practicality score for student use of the web-based learning evaluation is 87.42, categorized as practical. Meanwhile, the practicality score obtained by the administration of Universitas Negeri Padang is 89.33, also categorized as practical.

ACKNOWLEDGEMENT
The author would like to express gratitude to Universitas Negeri Padang for the smooth access to the data obtained during the research. The author also extends thanks to the lecturers of the Graduate Program in Educational Technology and Vocational Education for their guidance and direction, which enabled this research to proceed smoothly.

REFERENCES


