Classification of Non-Civil Servant Performance Appraisal Using Naïve Bayes Classifier Algorithm

Sofia Dewi
Information System Department, Ma’soem University, Indonesia

ABSTRACT
Employee performance assessment is a way to measure the level of employee productivity. In the process of assessing the performance of Non-Civil Servants (non-PNS) employees at the Regional Technical Implementation Unit of Education and Training of Cooperatives and Entrepreneurs (UPTD P3W) at this time, it is required to classify data based on several factors to find out whether the employee fits into the eligible category or not as the best employee to become a civil servant (PNS) candidate. The purpose of this research is to make it easier to determine the classification of the performance assessment of non-PNS employees at UPTD P3W using the Naïve Bayes Classifier Algorithm and to determine the level of accuracy in the classification of the performance assessment. In this study, the authors used 498 data as training data and 105 data as testing data for manual testing in Excel and for testing using RapidMiner tools. Based on the analysis in the study, the result of the predictions determines the best employees to become candidates for civil servants quickly and accurately, while from the tests performed by comparing training data and with data testing using RapidMiner tools, the accuracy rate is 84.76%.

Keywords: Classification, Naïve Bayes Classifier, Performance Assessment, Rapid Miner

Corresponding Author:
Sofia Dewi
Department of Information Systems, Ma’soem University, Indonesia
Jl. Raya Cipacing No.22, Cipacing, Kec. Jatinangor, Kabupaten Sumedang, Jawa Barat 45363
Email: sofiadewi.job@gmail.com

1. INTRODUCTION
In a company, both government and private companies, employees have a very important role in the company [1], [2]. In this very important role, employee performance needs to be managed properly [3], [4]. Employee performance needs to be measured to determine the maximum level of employee professionalism that has been achieved. Employee performance assessments like this have been performed by various organizations in the past.

Employee performance assessment is an effort to measure the level of employee productivity [1], [5]–[7]. Criteria that greatly influence performance assessment include length of service, age, gender, division, status, education, type of work, work behavior, personality aspects, technical work aspects, non-technical aspects, and goal attainment [8], [9]. The assessment standard that is applied is the attitude of employees and superiors when facing this assessment.

The Regional Technical Implementation Unit of Education and Training of Cooperatives and Entrepreneurs (UPTD P3W) is a government institution under the auspices of the Office of Cooperatives and Small Enterprises (KUK) - West Java Province [10], [11], located in Jl. Soekarno-Hatta No.708, Babakan Penghulu, Cinambo, Bandung City, West Java 40924. There are 55 employees at UPTD P3W, 24 civil servants (PNS), and 31 non-civil servants (non-PNS). 31 non-PNS employees are aged between 25 years and over 50 years with a working period of between 3 months and 14 years. Then, they will be classified as civil servant candidates, with a total of 498 data taken for data testing of 105 data. The problems in this research are:
a) In the process of assessing the performance of non-PNS employees currently at the UPTD, they still use manuals and already have assessment indicators and then it is necessary to classify data based on factors to determine whether the employee fits into the appropriate “eligible” or inappropriate “not eligible” category as a candidate for civil servants.

b) To find out the level of accuracy in classifying employee performance assessments.

2. METHOD
The method used is the Naïve Bayes Classifier. Naïve Bayes Classifier is one of the statistical classifiers [12]–[17], where this classifier can predict the probability of class membership of a data tuple that will enter a certain class according to probability calculations. The data mining method for classifying employee data is performed in several stages. The flow of the classification process can be seen in figure 1:

![Figure 1. Process of Naïve Bayes Classifier Method](image)

The flow of data mining using the Naïve Bayes Classifier method is shown in figure 2, which consists of several stages, including:

2.1. Data Collection
The data used in this research is using primary data. Primary data is the source of research data obtained directly from the source in the form of interviews. The data collected is data on daily reports on the activities of non-PNS employees at UPTD P3W - West Java, with a total of 498 data.

2.2. Preliminary Data Processing
Data preparation is taking data identified in the previous stage or preparing it for analysis using data mining methods. This is expected to be compared with other stages in CRISP-DM [18]–[20].

a) First, determining the data to be processed from the data that has been obtained, where not all data will be processed because the research to be carried out has limitations on the data used.

b) Second, handling missing value data. The missing value is incomplete data because the attribute is not recorded or the attribute is not owned. Handling of missing values is performed by deleting empty records.

c) Third, determining the attributes that will be used from the first stage.

2.3. The Proposed Method
After performing all stages of data processing, training data will then be generated. Training data is the data used to perform the calculation process [21], [22] using the Naïve Bayes Classifier method.

2.4. Final Testing/Validation
The test results will be validated and then evaluated using the RapidMiner software. RapidMiner is open-source software. RapidMiner is a solution for analyzing data mining, text mining, and predictive analysis. RapidMiner has approximately 500 data mining operators, including input, output, data preprocessing, and visualization operators.
3. RESULTS AND DISCUSSION

3.1. RESULTS

3.1.1. Initial Data

Data processing requires supporting data to determine whether the employee will be eligible or not as a candidate for civil servant employees. This data is used as a reference to determine the assessment of each employee, including:

a) Performance Assessment Support
b) Work Behavior
c) Personality Aspects
d) Technical Work Aspects
e) Non-Technical Aspects
f) Goal Attainment

Examples of data used in this study are data on non-civil servants at UPTD P3W West Java as shown in Figure 3.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Initial Period</th>
<th>Data</th>
<th>Flag</th>
<th>in [AM]</th>
<th>Out [PM]</th>
<th>Range</th>
<th>Activity</th>
<th>Information</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>06/05/2023</td>
<td>Monday</td>
<td>2:30</td>
<td>4:00</td>
<td>2</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>2</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>06/05/2023</td>
<td>Tuesday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>3</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>06/05/2023</td>
<td>Wednesday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>4</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>06/05/2023</td>
<td>Thursday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>5</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>06/05/2023</td>
<td>Friday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>6</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>20/06/2023</td>
<td>Monday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>7</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>20/06/2023</td>
<td>Tuesday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>8</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>20/06/2023</td>
<td>Wednesday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>9</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>20/06/2023</td>
<td>Thursday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>10</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>20/06/2023</td>
<td>Friday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>11</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Monday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>12</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Tuesday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>13</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Wednesday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>14</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Thursday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>15</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Friday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>16</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Monday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>17</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Tuesday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>18</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Wednesday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>19</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Thursday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>20</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Friday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>21</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Monday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>22</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Tuesday</td>
<td>2:30</td>
<td>5:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>23</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Wednesday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
<tr>
<td>24</td>
<td>Mh. Fstly</td>
<td>Technical Staff</td>
<td>05/07/2023</td>
<td>Thursday</td>
<td>2:30</td>
<td>6:00</td>
<td>3</td>
<td>Std</td>
<td>approved</td>
<td>applicable</td>
</tr>
</tbody>
</table>
3.1.2. Cleaning Data

Cleaning data is performed to remove noise and inconsistent data or irrelevant data, as shown in Figure 4.

3.1.3. Data Collection

Data collection is conducted to combine data from other sources, as shown in Figure 5.

3.1.4. Training Data

This training data uses data from a combined data of 498 data. Figure 6 below is a sample of 26 data.

---

**Figure 4. Cleaning Data**

**Figure 5. Data Collection**

**Figure 6. Training Data**
3.1.5. Testing Data

This testing data uses 105 data. Figure 7 below is the sample of 30 data.

3.2. Calculation of Naïve Bayes

\[
P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)
\]

\[
P(C_i) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (2)
\]

Where [23]–[25]:

- X = Data with unknown classes
- Ci = The X data hypothesis is a specific class
- P(Ci|X) = is the posterior probability of class Ci given input features X (posteriori prob)
- P(Ci) = Probability of hypothesis Ci (prior prob)
- P(X|Ci) = is the likelihood of observing the input features X given class Ci
- P(X) = is the marginal likelihood of the input features X (tuple X has the same probability of entering any class, so the maximized is P(X|Ci)P(Ci))

The first stage of calculation to determine the best non-PNS employee to become the PNS candidate with the Naïve Bayes method is to find the probability of each class. In determining the best employee, 2 classes will be determined: the "Eligible" and "Not Eligible" categories. The calculation method is to find out how much data is in the "Eligible" class and how much is in the "Not Eligible" class, from the total training data, then divide it by the total data. The results of the calculations can be seen in the figure 8 below:

![Figure 8. Class Probability](image)

3.3. Manual Calculation of Naïve Bayes

a) Calculation result (Example Set)

Sample 6 of data. As shown in Table 1 below:

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Criteria</th>
<th>Prediction (Criteria)</th>
<th>Confidence (Not Eligible)</th>
<th>Confidence (Eligible)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not Eligible</td>
<td>Not Eligible</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>Not Eligible</td>
<td>Not Eligible</td>
<td>0.998</td>
<td>0.002</td>
</tr>
</tbody>
</table>

![Table 1. Calculation Result (Example Set)](image)
b) The calculation result of Accuracy
From the results of Figure 4.11, it can be seen that the level of accuracy of the data used is 84.76% and the class precision "Eligible" is 84.72%, and "Not Eligible" is 84.85%. As for the class recall itself, for the "Eligible" class it is 92.42% and for the "Not Eligible" class it is 71.79%. In general, precision, recall, and accuracy can be formulated as follows:

i. For “Eligible” class
\[
\text{Precision} = \frac{61}{61+11} = \frac{61}{72} = 0.8472 = 84.72% \\
\text{Recall} = \frac{61}{61+5} = \frac{61}{66} = 0.9242 = 92.42% \\
\]

ii. For “Not Eligible” class
\[
\text{Precision} = \frac{28}{28+5} = \frac{28}{33} = 0.8485 = 84.85% \\
\text{Recall} = \frac{28}{28+11} = \frac{28}{39} = 0.7179 = 71.79% \\
\text{Accuracy} = \frac{(61+28)}{(61+5+11+28)} = \frac{89}{105} = 0.8476 = 84.76% \\
\text{Error} = \frac{(5+11)}{(61+5+11+28)} = \frac{16}{105} = 0.1524 = 15.24% \\
\]

3.4. Comparison of Accuracy
a) Comparison of accuracy
A comparison of overall accuracy with validation accuracy using Cross Validation is shown in Table 2.

<table>
<thead>
<tr>
<th>Accuracy %</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>84.76%</td>
<td>84.72%</td>
<td>92.42%</td>
</tr>
</tbody>
</table>

Cross Validation

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/- 14.13%</td>
<td>+/-13.97%</td>
<td>+/- 17.14%</td>
<td></td>
</tr>
</tbody>
</table>

b) Percentage of “Eligible” and “Not Eligible”
The percentage of Eligible and Not Eligible obtained from the results of testing 150 data with the Naïve Bayes Classifier, is shown in table 3.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Eligible</th>
<th>Not Eligible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>63%</td>
<td>37%</td>
</tr>
</tbody>
</table>

4. CONCLUSION
Based on the results of research that has been conducted on the classification of non-PNS performance assessments using the Naïve Bayes Classifier Algorithm method at UPTD P3W - West Java Province, it can be concluded as follows:

a) This performance assessment has 15 indicators, such as gender, age, status, education, employment, length of service, division, work behavior, personality aspects, technical work aspects, non-technical aspects, goal attainment, activities, information, and criteria. The classification process using the Naïve Bayes method uses training data to generate probabilities for each criterion in different classes. These probability values can be optimized to determine the best employee to become a civil servant (PNS) candidate at UPTD P3W - West Java Province.

b) Based on the results of the analysis in the research, it is found that the prediction results determine the best employees become candidates for civil servants quickly and accurately. From the tests performed by comparing training data with testing data using RapidMiner supporting tools, it is obtained that the accuracy rate is 84.76%.

REFERENCES

Classification of Non-Civil Servant Performance Appraisal ... (Sofia Dewi) - 80


[21] D. H. Kamagi and S. Hansun, “Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Classification of Non-Civil Servant Performance Appraisal ... (Sofia Dewi) - 81
Tingkat Kelulusan Mahasiswa,” *J. Ultim.*, vol. 6, no. 1, pp. 15–20, Jun. 2014.


