KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol. 20 (2)(2023), 126-134 p-ISSN: 1693-7554, e-ISSN:2654-3990

C4.5 Algorithm Implementation to Predict Student Satisfaction Level of Lecturer's Performance in the Covid-19 Pandemic

Juan R. M. Ledoh[1*], **Ferdinandus E. Andreas**[2], **Emerensye S. Y. Pandie**[3], **Clarissa E. Amos Pah**[4]

[1,2,3,4] Computer Science Study Program, Faculty of Science and Engineering, Universitas Nusa Cendana, Kupang, East Nusa Tenggara, Indonesia

---

Abstract

Abstract Implementation of education during the emergency period of Covid-19 in Higher Education was carried out at home through online/distance learning. The lecturer is one of the key holders of success in the learning process. Lecturer performance is a main factor needed to improve education and service quality in online learning. In this study, the authors implemented the C4.5 algorithm using RapidMiner 9.10 app to predict student satisfaction with lecturer performance during the Covid-19 pandemic. The data in this study were obtained from a questionnaire distributed to active students in the Computer Science Study Program (class of 2016 - 2021) at the University of Nusa Cendana with 942 records. The attributes used in this study were the lecturer's Age, Gender, Suitability of Learning Media (SLM), and the competencies of Pedagogic Competence (PeC), Professional Competence (PrC), Personal Competence (PsC), and social competence (SC), with the level of student satisfaction as the target class divided into two, namely *Satisfied* and *Dissatisfied*. The data-sets is processed using RapidMiner and produces 11 decision rules which show that the attribute PeC has the most significant influence on the level of student satisfaction with lecturer performance during the Covid-19 pandemic and the test results of the decision tree model using cross-validation. The test results show that the C4.5 algorithm has a good performance in predicting levels of student satisfaction with an accuracy rate of 94.8%, precision for the prediction class *Dissatisfied* and *Satisfied* of 92.23% and 95.52%, and recall of the actual *Dissatisfied* and *Satisfied* classes of 85.2% and 97.77%.

Keywords: *Student Satisfaction; Lecturer Performance; C4.5 Algorithm; RapidMiner*

---

1. Introduction

The spread of Covid-19 in Indonesia in early 2020 significantly impacted various aspects, including the world of education. To prevent the spread of Covid-19, on March 24 2020, the government issued a Circular Letter of the Minister of Education and Culture of the Republic of Indonesia Number 4 of 2020 About The Implementation of Education Policies during the Emergency Period of the Spread of Covid-19 which explains that the learning process in each institution or educational institutions implemented at home distance learning through online learning.

*Corresponding author. *E-mail address: juanledoh@staf.undana.ac.id*
Received: 30 June 2023, Accepted: 07 July 2023 and available online 26 July 2023
http://dx.doi.org/10.33751/komputasi.v20i2.8284

In the implementation of education during the Covid-19 pandemic, the quality of lecturer performance is a central factor needed to improve the quality of education and service quality in the online learning process in Higher Education, where if the quality of lecturer performance is very good, it can facilitate the delivery of knowledge and technology so that successful teaching and learning process in Higher Education during the Covid-19 pandemic can be achieved.

Regarding the implementation of education during the Covid-19 pandemic, the Computer Science Study Program at Nusa Cendana University began implementing online learning in March 2020. In the implementation of online learning, various obstacles emerged such as limited internet access, reduced interaction with lecturers, reduced student understanding of learning material, and lack of supervision in learning. In addition, there were also obstacles presented by students related to lecturer performance in providing online learning including the material provided was not in-depth, the lecturer's explanations were difficult to understand, and the material content provided in the LMS (Learning Management System) or E-Learning which is still minimal in supporting student understanding.

To increase the satisfaction of students, lecturers must continue to improve performance/service oriented to student needs in online learning for student satisfaction and comfort during online learning. To support the improvisation of teaching, we must analyze the level of student satisfaction with the performance of lecturers during online learning during the Covid-19 pandemic. In this study, researchers will use the C4.5 algorithm with the help of the RapidMiner 9.10 application. The C4.5 algorithm is one of the data mining algorithms that produce decision trees that are easy to interpret, have an acceptable level of accuracy, and are efficient in handling discrete and numeric type attributes/variables [1] [2]. In the process, researchers refer to several previous studies that can support research, such as (1) Yuliana and Pratomo in predicting student satisfaction with the performance of lecturers at Politeknik Tedc Bandung. This research resulted in predicting student satisfaction with lecturer performance using the C.45 algorithm very well, which is shown by the accuracy rate of 94.62% [3], (2) Hidayah and Rozi in knowing the performance of the best employees at Universitas Mercu Buana Yogyakarta with the results of the application of the C4.5 algorithm has an accuracy level that is included in the excellent classification of 85.52% [4], and (3) Permana in predicting student satisfaction with the level of service at the Faculty of Engineering, Hamzanwadi University by utilizing the C4.5 algorithm and obtained the C4.5 algorithm can predict satisfaction with the highest accuracy value of 81.07% [5]. The studies mentioned show that the C4.5 algorithm produces a decision tree with an average accuracy of more than 80% in predicting the satisfaction or performance of lecturers. So, in this this study may help Computer Science institutions and lecturers to find out an overview of the factors that affect student satisfaction with lecturer performance during the Covid-19 pandemic.

2. Methods

2.1. Data Mining

Data mining is an activity of extracting or mining and analyzing large amounts of data to find patterns and rules that are useful for development [1][6]. Data mining is also a part of the entire process in Knowledge Discovery in Database (KDD). In this section certain algorithms or methods will be used to perform calculations in order to produce patterns from the processed data, then new knowledge will be found in the resulting patterns [7].

2.2. Classification

One technique commonly used in data mining is classification. Several algorithms that can be used for classification are C4.5 (Decision Tree), Naïve Bayes, Neural Network, K-Nearest Neighbor, and others. Classification is the process of placing an object into a predefined category/class based on a certain model. In the classification process there are two stages that must be passed, namely the learning and testing stages. In the learning stage some of the data whose data class is known (data training) is used to form an approximate model, while in the testing stage the forecast model that has been formed is tested with some other data (data testing) to determine the accuracy of the model, where if the accuracy can be accepted, this model is used to predict unknown data classes [1].

2.3. Decision Tree

The most popular data mining classification model is the decision tree, which is used as a data reasoning procedure to find solutions to the problems being analyzed by turning large facts into decision trees that represent rules [8][9]. The working concept of this tree structure is to transform data (tables) into a decision tree model, then convert the tree into a rule or decision rule, and then these rules are simplified. The working concept of this tree structure can be seen in Figure 1.

Figure 1. Decision Tree Concepts

2.4. C4.5 Algorithm

C4.5 algorithm is a data mining algorithm that is used to create a decision tree model. The C4.5 algorithm is a development of the ID3 algorithm, where the working principle of the C4.5 algorithm is similar to the ID3 algorithm, the things that distinguish the C4.5 algorithm from ID3 include being able to handle attributes with discrete or continuous types, being able to handle attributes that have missing values, and able to prevent data noise [10]. The advantages of the C4.5 algorithm are that it can process numeric and discrete data, can handle missing attribute values, produces decision rules that are easy to interpret, and its performance is one of the fastest compared to other algorithms[11].
The stages of the C4.5 algorithm in making a decision tree model are as follows [12]:

1. Data-sets Prepare.
2. Calculating the value of information gain.
   To select an attribute as the root/root node/node 1, it is based on the highest information gain value of the existing attributes. The formula for calculating the information gain for each attribute can be seen in Eq. 1.

   $$InformationGain(A) = Entropy(S) - Entropy_A(S) \tag{1}$$

   Information:
   A = Attribute/variable
   S = case set
   Meanwhile, to calculate the value of $Entropy(S)$ or total entropy using Eq. 2 and to calculate the entropy value for each attribute or $Entropy_A(S)$ you can use Eq. 3.

   $$Entropy(S) = \sum_{i=1}^{n} -p_i \, log_2 \, p_i \tag{2}$$

   Information:
   n = Number of partitions S
   p_i = Number of proportions for class i

   $$Entropy_A(S) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i) \tag{3}$$

   Information:
   n = Number of partitions attribute A
   $|S_i|$ = Number of cases on i-th partition
   $|S|$ = Number of cases in set of S

3. Selecting the attribute as the root node.
4. Repeat step number 2 and step number 3 until every branch is fulfilled.
5. The decision tree partition process will stop when all branches in node have the same class.

2.5. Lecture's Performance

Performance is the work achieved by a workforce in carrying out their duties in accordance with the responsibilities given to them [13]. Thus, lecturer's performance is the work achieved by a lecturer in carrying out his duties and responsibilities as an educator or human resource in higher education that transforms, develops, and disseminates science, technology, and art through education, research, and community service. Lecturer performance indicators are largely determined by their competence, where competence itself is a set of knowledge, skills, and behaviors that must be owned, lived, and mastered by lecturers in carrying out the duties assigned to them. These competencies include: pedagogical, professional, personality, and social [14].

2.6. Students' Satisfaction

Student satisfaction means feeling satisfied and happy with the quality of education, especially the quality of service in the learning process at tertiary institutions provided by lecturers, where the learning satisfaction felt by students indicates that these students enjoy the learning process provided by lecturers [15].

2.7. Research Methods

This research uses data mining stages as shown in Figure 2. The stages consist of data collection, data selection, data preprocessing, data transformation, data mining method implementation, and results interpretation.

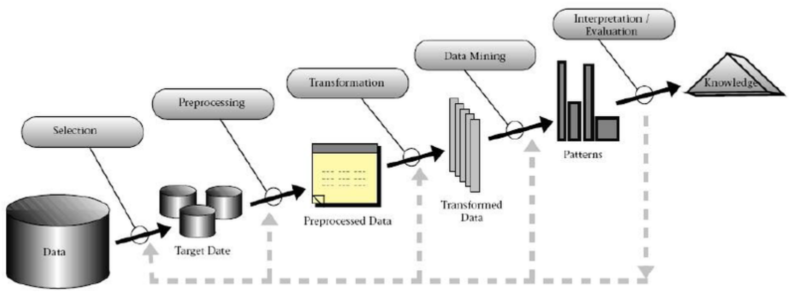

Figure 2. Data Mining Stages [7]

1. Data Collection
   The data collected are the primary data that was obtained directly from respondents' responses to the satisfaction questionnaires of lecturer performance.
2. Data Selection
   The collection data is then selected according to the research requirements. The output of this stage is some attributes and class that will be used for data mining classification.
3. Data Preprocessing
   All the data rows are through the cleansing process by removing the redundant records and the records that have empty fields.
4. Data Transformation
   The clean data collection is then transformed to produce the final data-set that will be used in the data mining method implementation.

5. C4.5 Algorithm Implementation
   In this stage, the C4.5 algorithm is run with RapidMiner tool. There are 5 used operators to construct a decision tree model and produce the data mining performance with 10-Fold Cross Validation.

3. Result and Discussion

This section explains the output of each data mining stage from the data preparation process (data collection, data selection, data preprocessing, and data transformation), C4.5 Algorithm Implementation, until the evaluation process.

3.1. Data Preparation

All the data collection are gained from the respondents' response to the lecture performance questionnaire. The involved respondents were active students from class 2016-2022 of the Computer Science Study Program, Nusa Cendana University. There are 942 clean data records that have been through the data collection, data selection, data preprocessing, data transformation. In data selection stages, researchers determined the used attributes and class. There are 7 selected attributes and 1 class with 2 class values that used for classification. The attributes are Age, Gender, Suitability of Learning Media (SLM), Pedagogic Competence (PeC), Professional Competence (PrC), Personality Competence (PsC), and Social Competence (SC). The class is Satisfaction Level that has 2 values, namely *Incompetent* and *Competent*. Table 1 shows the attribute list and Table 2 shows the class. In this research, the data preprocessing is done with no changes on the data-set

Table 1. Attribute List

No.	Attribute	Attribute Value
1.	Age	25-35 yo 36-46 yo >46 yo
2.	Gender	Female (F) Male (M)
3.	Suitability of Learning Media (SLM)	Yes No
4.	Pedagogic Competence (PeC)	Incompetent Competent
5.	Professional Competence (PrC)	Incompetent Competent
6.	Personality Competence (PsC)	Incompetent Competent
7.	Social Competence (SC)	Incompetent Competent

Table 2. Class

Class	Class Value
Satisfaction Level (SL)	*Dissatisfied* *Satisfied*

because there is no record that has empty fields or redundant records. This data-set also does not need transformation so the data that has been through the preprocessing is the final data-set as shown in Table 3.

Table 3. Final Data-set for Data Mining Implementation

No.	Gender	Age	SLM	PeC	PrC	PsC	SC	SL
1.	F	36 - 46 yo	No	Incomp.	Incomp.	Comp.	Incomp.	*Dissatisfied*
2.	F	25 - 35 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
3.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
4.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
5.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
6.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
7.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
8.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
9.	M	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
10.	F	36 - 46 yo	Yes	Comp.	Comp.	Comp.	Comp.	*Satisfied*
...	...	...	...	...	...	...	...	...
942.	M	$>$ 46 yo	No	Incomp.	Incomp.	Incomp.	Incomp.	*Dissatisfied*

3.2. C4.5 Algorithm Implementation

To apply the C4.5 algorithm, researches use the RapidMiner tool with apply some operators. The operators are Read Excel, Cross-Validation, Decision Tree, Apply Model, and Performance. Read Excel is used to import raw data from a document with Microsoft Excel format and determine which column as the attributes or the class. Decision Tree is used to construct and display the decision tree that represents the decision rules. Cross Validation is used to evaluate the decision tree model with apply the K-Fold Cross Validation. Apply Model is an operator that can run inside the cross-validation operator that is used to apply the decision tree model to the testing data-set to produce the prediction. Performance uses the prediction data to build a confusion matrix that will be used to calculate accuracy, precision, and recall. The decision tree result is shown in Figure 3.

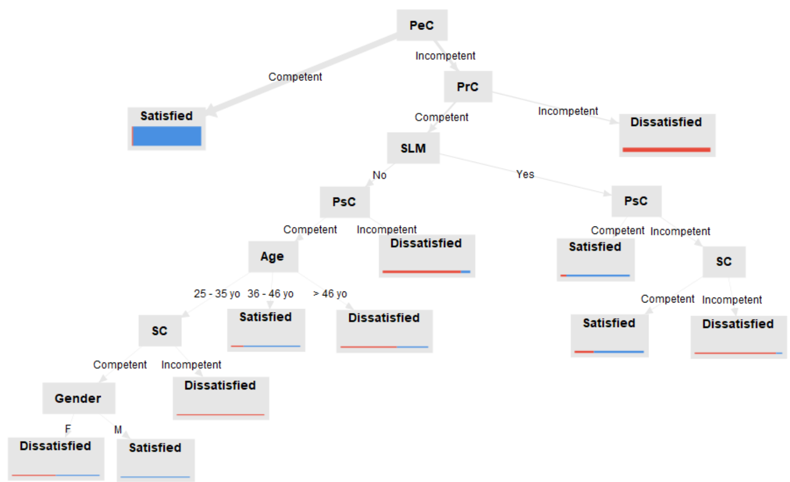

Figure 3. C4.5 Decision Tree Model

As illustrated in Figure 3, the Decision Tree that is formed has 2 points of discussion, namely:

1. Most significant attributes to less significant attributes to the data.
   (a) The PeC attribute is the attribute that has the most significant effect on class classification. It can be seen that the PeC attribute is the root attribute which is used as a reference in the Decision Tree. If the lecturer *Competent* to the PeC attribute, then the classification results will go to the *Satisfied* class. It's another case if the lecturer

Incompetent the PeC attributes, then the next influential attribute will be traced from the entire data.

(b) The PrC attribute is the second significant attribute of the data. The Decision Tree shows tracing after the PeC attribute (the most influential attribute) going to the root of the PrC attribute. If the lecturer is *Incompetent* to the PrC attribute, then the classification results will go to the *Dissatisfied* Class. Different things are shown if the lecturer *Competent* to the PrC attribute, then the next influential attribute will be traced.

(c) The SLM attribute is an attribute that is influential after PeC and PrC are in accordance with the search from the root of the Decision Tree. The SLM attribute will then explore the next attribute even though this attribute contains data that is not match (*No*) or match (*Yes*) with the learning media owned by the lecturer.

(d) The PsC attribute or Personality Competence is the next attribute to be traced after the PeC, PrC, and SLM attributes so that it can be considered that the PsC attribute is the fourth attribute that influences the data. The PsC attributes that are traced are divided based on the decision value of the SLM attributes. If the PsC attribute receives a decision value of *No*, then the PsC will result in a class classification of *Dissatisfied* if and only if the PsC attribute value is *Incompetent* and will move to the next significant attribute if *Competent*. In contrast, if the PsC receives a decision value of *Yes* from the SLM attribute, then the PsC will produce a *Satisfied* class if and only if the PsC attribute value is *Competent* and will move to the next significant attribute if *Incompetent*.

(e) The Age and SC are the next attributes that are triggered if traced from the influential attributes of PeC, PrC, SLM, and PsC. It can be said that the Age and SC is an attribute that influences the data but has the same level. However, when viewed from the Decision Tree, the Age attribute will trace to other attributes while the SC attribute will produce a class regardless of the value of the SC attribute. So, significantly between the two attributes, the SC attribute has more influence on the data than the Age attribute.

(f) The Age attribute is the next influential attribute after the PeC, PrC, SLM, PsC, and SC attributes. The Age attribute will produce a *Satisfied* class if the Age is between "36-46 yo" and produce a *Dissatisfied* class if the Age is more than 46 years old ("> 46 yo"). The Age attribute will browse to other attributes when the attribute value is "25-35 yo".

(g) The Gender attribute is the last attribute that is triggered after tracing the significant attributes of PeC, PrC, SLM, PsC, SC, and Age. It can be said that this attribute is a less significant attribute to the data. It can be seen in the Decision Tree that the Gender attribute is the attribute with the bottom level of the Decision tree.

So if the attributes are based on influence on the data from the most influential to the least influential, then the attributes will be sorted into PeC, PrC, SLM, PsC, SC, Age, and Gender.

2. Based on the Decision Tree, decision rules can be formed as shown in Table 4.

accuracy: 94.80% +/- 2.98% (micro average: 94.80%)

	true Dissatisfied	true Satisfied	class precision
pred. Dissatisfied	190	16	92.23%
pred. Satisfied	33	703	95.52%
class recall	85.20%	97.77%	

Figure 4. Confusion Matrix

Table 4. Decision Rules Based on the Decision Tree

Rules	Decision Rules	Class
1.	IF PeC = *Competent*	Satisfied
2.	IF PeC = *Incompetent* AND PrC = *Incompetent*	*Dissatisfied*
3.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *Yes* AND PsC = *Competent*	*Satisfied*
4.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *Yes* AND PsC = Incompetent AND SC = Competent	*Satisfied*
5.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *Yes* AND PsC = *Incompetent* AND SC = *Incompetent*	*Dissatisfied*
6.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *No* AND PsC = *Incompetent*	*Dissatisfied*
7.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *No* AND PsC = *Competent* AND Age = >46 yo	*Dissatisfied*
8.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *No* AND PsC = *Competent* AND Age = 36 - 46 yo	*Satisfied*
9.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *No* AND PsC = *Competent* AND Age = 25 - 35 yo AND SC = *Incompetent*	*Satisfied*
10.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *No* AND PsC = *Competent* AND Age = 25 - 35 yo AND SC = *Competent* AND Gender = *M*	*Satisfied*
11.	IF PeC = *Incompetent* AND PrC = *Competent* AND SLM = *No* AND PsC = *Competent* AND Age = 25 - 35 yo AND SC = *Competent* AND Gender = *F*	*Dissatisfied*

Performance Based on Figure 4, the accuracy value obtained from testing using 10-fold cross validation on the decision tree model is 94.80%, precision for the prediction class Dissatisfied and Satisfied of 92.23% and 95.52%, and Recall of the actual *Dissatisfied* and *Satisfied* classes of 85.2% and 97.77%. In addition, the author also analyzes the decision rules generated by the decision tree model to find the most significant attributes (discussion point 1 on Figure 3) in influencing student satisfaction with lecturer performance in the Covid-19 pandemic, the results obtained show that the Pedagogic Competence (PeC) is the attribute that most affects student satisfaction with lecturer performance in the online learning process during the Covid-19 pandemic.

4. Conclusion

Algorithm C4.5 has a good performance in classifying the level of student satisfaction with lecturer performance during the Covid-19 pandemic at Computer Science Program Study Nusa Cendana University with an accuracy rate of 94.8%, precision for the prediction class *Dissatisfied* and *Satisfied* of 92.23% and 95.52%, and recall of the actual *Dissatisfied* and *Satisfied* classes of 85.2% and 97.77%. All attributes used in this study (7 attributes) were triggered as nodes in the decision tree which showed that all attributes had an effect on lecturer performance. The PeC attribute is the attribute that has the most influence on lecturer performance, followed by the attributes of PrC, SLM, PsC, SC, Age, and the less influential attribute is Gender.

References

[1] J. Han and M. Kamber, "Data mining: Concepts and Techniques," Morgan kaufmann, vol. 340, pp. 94 104–3205, 2006.

[2] C. E. Amos Pah and D. N. Utama, "Decision Support Model for Employee Recruitment using Data Mining Classification," International Journal, vol. 8, no. 5, pp. 1511–1516, 2020.

[3] A. Yuliana and D. B. Pratomo, “Algoritma Decision Tree (C4. 5) untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung,” in Prosiding SemNas InoTek (Seminar Nasional Inovasi Teknologi), vol. 1, no. 1, 2017, pp. 377–384.

[4] A. N. Z. Hidayah and A. F. Rozi, “Penerapan data mining dalam menentukan kinerja karyawan terbaik dengan menggunakan metode algoritma C4. 5 (studi kasus: Universitas Mercu Buana Yogyakarta),” Journal Of Information System And Artificial Intelligence, vol. 1, no. 2, pp.117–127, 2021.

[5] B. A. C. Permana, “Prediksi Kepuasan Mahasiswa Terhadap Tingkat Pelayanan Menggunakan Algoritma C4. 5 (Decision Tree)(studi kasus: Fakultas teknik Universitas Hamzanwadi),” Infotek: Jurnal Informatika dan Teknologi, vol. 1, no. 1, pp. 32–39, 2018.

[6] C. C. Aggarwal, Data mining: the textbook. Springer, 2015.

[7] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth et al., “Knowledge Discovery and Data Mining: Towards a Unifying Framework,” in KDD, vol. 96, 1996, pp. 82–88.

[8] D. Jollyta, W. Ramdhan, and M. Zarlis, Konsep Data Mining Dan Penerapan. Deepublish, 2020.

[9] Suyanto, Data Mining untuk Klasifikasi dan Klasterisasi Data. Bandung: Informatika, 2017.

[10] E. Buulolo, Data Mining Untuk Perguruan Tinggi. Deepublish, 2020.

[11] R. T. Vulandari, Data Mining Teori dan Aplikasi Rapidminer. Yogyakarta: Gava Media, 2017.

[12] L. Swastina, “Penerapan Algoritma C4. 5 untuk Penentuan Jurusan Mahasiswa,” Jurnal Gema Aktualita, 2013.

[13] V. Rivai, Manajemen Sumber Daya Manusia untuk Perusahaan. Rajawali Pers, 2016. http://www.ieomsociety.org/harare2020/papers/434.pdf

[14] P. Peraturan, “Peraturan Pemerintah Republik Indonesia nomor 37 Tahun 2009 Tentang Dosen,” 2009.

[15] Y. Dafroyati, “Kepuasan Mahasiswa Terhadap Kinerja Dosen dalam Pembelajaran pada Jurusan dan Program studi di Lingkup Poltekkes Kemenkes Kupang Tahun 2014,” Jurnal Info Kesehatan, vol. 14, no. 1, pp. 1135–1148, 2016.