ISSN No:-2456-2165

Lecturer Research Performance Model Evaluation using Machine Learning Approach

Ahmad Sanmorino
Faculty of Computer Science
Universitas Indo Global Mandiri
Palembang, Indonesia

Abstract:- Through this study, the author evaluates the lecturer's research performance model built on previous research. This model consists of seven independent variables and one dependent variable. The seven independent variables that construct the model are Scientific Article, H-Score, College Type, Journal Cluster, Research Grant, Research Collaboration, Research Interest, while the dependent variable is research performance. Based on the results of the evaluation using the machine learning approach, a good accuracy score was obtained for each classifier, for Random Forest at 93 percent, Multi-layer Perceptron at 90 percent, Decision Tree at 97 percent, and Linear Discriminant Analysis at 93 percent. The results of this evaluation show that the proposed research performance model of the lecturer meets the author's expectations and is relevant to the conditions of higher learning institutions.

Keywords:- *Research Performance Model; Lecturer; Evaluation; Machine Learning.*

I. INTRODUCTION

This study is a continuation of the previous related study. In the previous study, the author has built a lecturer research performance model. In this study, an evaluation of the proposed model was carried out. Evaluation of the model involves four machine learning classifiers. The model building is expected to be a reference for research managers in assessing and improving the research performance of lecturers at higher learning institutions. The research performance which is judged from the quality and quantity of research output is still not optimal, with the model that the author proposes to be an alternative solution for the problem.

II. RELATED STUDY

In previous study [1][2], the author has built a lecturer research performance model shown in Fig. 1:

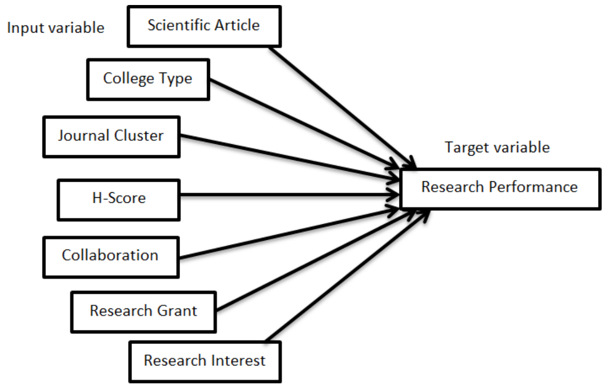

Fig. 1. Conceptual Model

This model consists of seven independent variables and one dependent variable (See Fig. 1). The author has not conducted feature selection and evaluation of the proposed model. In this study, only evaluation of the model is carried out, for feature selection will be carried out in other studies.

Other related studies published by Wichian et al. [3], investigated the factors that influence research productivity in public universities. Research Management, Research Funding, Communication, Networking-Teamwork, Age, Academic Position, Thinking, Research Mind, Volition-Control, International Meeting, Research Skill-Techniques, Institutional Policy, and Library Expenditure are the variables used in this study. The empirical data support the research productivity model (Chi-Square at 80.007). The Chi-Square function is used to calculate the degree of relationship between variables [4]. Back Propagation Neural Networks are used by the authors to analyze the factors that influence research performance [5].

III. METHODOLOGY

The steps that the author uses as a guide in conducting this study are shown in research design in Fig. 2:

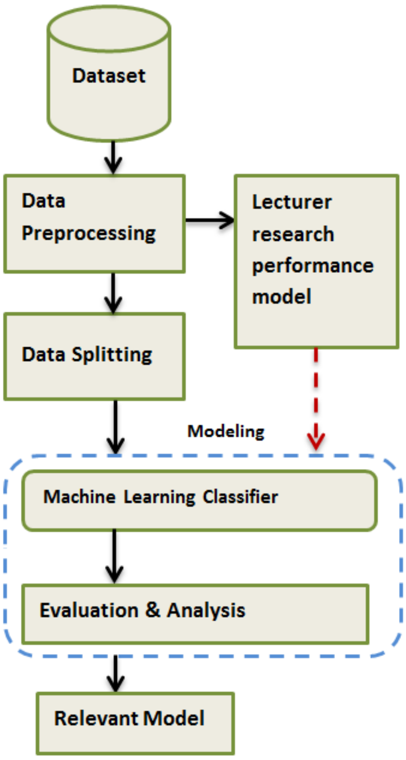

Fig. 2. Research Design

This study begins with study literature publications related to lecturer research performance, followed by the definition of variables that are used as the proposed construct model. Then collect data based on the variables that have been determined, followed by the development of a lecturer research performance model. Determination of variables, data collection, and model development has been carried out by the authors in previous research [1][2], so that this study enters the next stage, the stage of evaluating the model that has been built. This evaluation phase involves four machine learning classifiers [6][7], Random Forest (RF) [8], Multi-layer Perceptron (MLP) [9][10], Decision Tree (DT) [11], and Linear Discriminant Analysis (LDA) [12]. The classification results are evaluated using several mechanisms, confusion-matrix [13][14], accuracy [15], precision, recall [16][17], f1-score [18], and AUC (area under the curve)[19][20]. In the last section, the author compares the results of the evaluation of each classifier, to find out whether the proposed model is relevant for use in higher learning institutions or vice versa.

IV. RESULT AND DISCUSSION

The evaluation results of the proposed model consist of confusion-matrix, accuracy score, precision, recall, f1-measure, misclassification rate, and others (See TABLE I – TABLE V). The comparison of True Positive, False Positive, True Negative, and False Negative scores are shown in TABLE I:

TABLE I. CONFUSION MATRIX REPORT

Classifier	True Positive	False Positive	True Negative	False Negative
RF	30 %	0 %	63.33 %	6.67 %
MLP	20 %	10 %	70 %	0 %
DT	26.67 %	0 %	70 %	3.33 %
LDA	23 %	0 %	73.67 %	3.33 %

Confusion-Matrix using Random Forest classifier, the number of correctly identified lecturers who did not meet the research performance target was 63.33 percent. 6.67 percent of lecturers were incorrectly identified as not meeting the research performance target. 30 percent of lecturers are correctly identified that they met the research performance target. 0 percent of lecturers were incorrectly identified as meeting the research performance target.

Confusion-Matrix using a Multi-layer Perceptron classifier, the number of correctly identified lecturers who did not meet the research performance target was 70 percent. 0 percent of lecturers were incorrectly identified as not meeting the research performance target. 20 percent of lecturers are correctly identified that they met the research performance target. 10 percent of lecturers were incorrectly identified as meeting the research performance target.

Confusion-Matrix using Decision Tree classifier, the number of correctly identified lecturers who did not meet the research performance target was 70 percent. 3.33 percent of lecturers were incorrectly identified as not meeting the research performance target. 26.67 percent of lecturers are correctly identified that they met the research performance target. 0 percent of lecturers were incorrectly identified as meeting the research performance target.

Confusion-Matrix using Linear Discriminant Analysis classifier, the number of correctly identified lecturers who did not meet the research performance target was 73.67 percent. 3.33 percent of lecturers were incorrectly identified as not meeting the research performance target. 23 percent of lecturers are correctly identified that they met the research performance target. 0 percent of lecturers were incorrectly identified as meeting the research performance target.

The results of the evaluation using the Random Forest (RF) method are shown in TABLE II:

TABLE II. RANDOM FOREST EVALUTION REPORT

	Precision	Recall	f1-Score
0	0.83	1.00	0.90
1	1.00	0.90	0.95
Accuracy			0.93
Macro avg	0.91	0.95	0.93
Weighted avg	0.95	0.93	0.94

Random Forest (RF) produced 83 percent of lecturers who did not meet the research performance target of all lecturers who were predicted to fail. Of all lecturers predicted to meet the research performance target, RF produces 100 percent who can actually meet it. In comparison to all lecturers who do not meet the research performance target, RF produces 100 percent of lecturers who are predicted not to meet the research performance target. In comparison to all lecturers who actually meet the research performance target, RF produces 90 percent of those who are predicted to meet it. RF generates a comparison of average precision and recall for lecturers who do not meet the 90 percent research performance target. RF generates a comparison of the average precision and recall for lecturers who meet the 95 percent research performance target (f-measure). RF produced 93 percent of lecturers who were correctly predicted to meet the research performance target but did not meet the overall lecturers' research performance target (accuracy). Furthermore, the evaluation using the Multi-layer Perceptron (MLP) method is shown in TABLE III:

TABLE III. MULTI-LAYER PERCEPTRON EVALUTION REPORT

	Precision	Recall	f1-Score
0	1.00	0.67	0.80
1	0.88	1.00	0.93
Accuracy			0.90
Macro avg	0.94	0.83	0.87
Weighted avg	0.91	0.90	0.89

Multi-layer Perceptron (MLP) produced 100 percent of lecturers who did not meet the research performance target of all lecturers who were predicted to fail. Of all lecturers predicted to meet the research performance target, MLP produces 88 percent who can actually meet it. In comparison to all lecturers who do not meet the research performance target, MLP produces 67 percent of lecturers who are predicted not to meet the research performance target. In comparison to all lecturers who actually meet the research performance target, MLP produces 100 percent of those who are predicted to meet it. MLP generates a comparison of average precision and recall for lecturers who do not meet the 80 percent research performance target. MLP generates a comparison of the average precision and recall for lecturers who meet the 93 percent research performance target (f-measure). MLP produced 90 percent of lecturers who were correctly predicted to meet the research performance target but did not meet the overall lecturers' research performance target (accuracy). Furthermore, the evaluation using the Decision Tree (DT) method is shown in TABLE IV:

TABLE IV. DECISION TREE EVALUTION REPORT

	Precision	Recall	f1-Score
0	0.89	1.00	0.94
1	1.00	0.95	0.98
Accuracy			0.97
Macro avg	0.94	0.98	0.96
Weighted avg	0.97	0.97	0.97

Decision Tree (DT) produced 89 percent of lecturers who did not meet the research performance target of all lecturers who were predicted to fail. Of all lecturers predicted to meet the research performance target, DT produces 100 percent who can actually meet it. In comparison to all lecturers who do not meet the research performance target, DT produces 100 percent of lecturers who are predicted not to meet the research performance target. In comparison to all lecturers who actually meet the research performance target, DT produces 95 percent of those who are predicted to meet it. DT generates a comparison of average precision and recall for lecturers who do not meet the 94 percent research performance target. DT generates a comparison of the average precision and recall for lecturers who meet the 98 percent research performance target (f-measure). DT produced 97 percent of lecturers who were correctly predicted to meet the research performance target but did not meet the overall lecturers' research performance target (accuracy). Furthermore, the evaluation using the Linear Discriminant Analysis (LDA) method is shown in TABLE V.

Linear Discriminant Analysis (LDA) produced 78 percent of lecturers who did not meet the research performance target of all lecturers who were predicted to fail. Of all lecturers predicted to meet the research performance target, LDA produces 100 percent who can actually meet it. In comparison to all lecturers who do not meet the research performance target, LDA produces 100 percent of lecturers who are predicted not to meet the research performance target.

TABLE V. LINEAR DISCRIMINANT ANALYSIS EVALUTION REPORT

	Precision	Recall	f1-Score
0	0.78	1.00	0.88
1	1.00	0.91	0.95
Accuracy			0.93
Macro avg	0.89	0.96	0.91
Weighted avg	0.95	0.93	0.94

In comparison to all lecturers who actually meet the research performance target, LDA produces 91 percent of those who are predicted to meet it. LDA generates a comparison of average precision and recall for lecturers who do not meet the 88 percent research performance target. LDA generates a comparison of the average precision and recall for lecturers who meet the 95 percent research performance target (f-measure). LDA produced 93 percent of lecturers who were correctly predicted to meet the research performance target but did not meet the overall lecturers' research performance target (accuracy).

The comparison of accuracy and misclassification rate for each classifier is shown in TABLE VI:

TABLE VI. THE COMPARISON OF ACCURACY AND MISCLASSIFICATION RATE

Classifier	Accuracy	AUC	Misclassification
RF	93 %	95 %	7 %
MLP	90 %	83 %	10 %
DT	97 %	98 %	3 %
LDA	93 %	96 %	7 %

The highest accuracy score is Decision Tree for 97 percent, followed by 93 percent for Random Forest and Linear Discriminant Analysis, the last, Multi-layer Perceptron at 90 percent. The goal of this evaluation is not to find the best accuracy score, but to see if the variables that comprise the model can pass the test phase with expected results. A good or relevant result is one with an accuracy score of more than 70%. The wider the area under the curve (AUC), the better the qualification results, the AUC of Decision Tree has the highest score compared to other classifiers.

V. CONCLUSION

The test results of the proposed lecturer research performance model obtained a high score for each classifier. Evaluation using confusion-matrix, accuracy, precision, recall, and f-measure shows good results. The accuracy score for Random Forest is 93 %, Multi-layer Perceptron is 90%, Decision Tree is 97 %, and Linear Discriminant Analysis is 93 %. The results of this evaluation show that the proposed model is relevant to real conditions in higher learning institutions. In future work, the author will add variables that construct the model, and perform testing with different combinations of machine learning classifiers.

REFERENCES

[1]. A. Sanmorino, Ermatita, Samsuryadi, and D. P. Rini, "Building Research Productivity Framework in Higher Education Institution," Int. Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 1–9, 2021.
[2]. A. Sanmorino, Ermatita, and Samsuryadi, "The preliminary results of the kms model with additional elements of gamification to optimize research output in a higher education institution," Int. J. Eng. Adv. Technol., vol. 8, no. 5, pp. 554–559, 2019.
[3]. S. Na Wichian, S. Wongwanich, and S. Bowarnkitiwong, "Factors affecting research productivity of faculty members in government universities: LISREL and Neural Network analyses," Kasetsart J. - Soc. Sci., vol. 30, no. 1, pp. 67–78, 2009.
[4]. K. Molugaram and G. S. Rao, Chi-Square Distribution. 2017.
[5]. F. Allegrini and A. C. Olivieri, Chemometrics and statistics | neural networks, 3rd ed. Elsevier Inc., 2019.
[6]. S. Theodoridis, Classification: a Tour of the Classics. 2020.
[7]. H. Anns, Basic classification concepts 13. 2018.
[8]. A. Urso et al., "Data Mining: Prediction Methods," pp. 1–18, 2018, doi: 10.1016/B978-0-12-809633-8.20462-7.
[9]. A. Urso, A. Fiannaca, M. La Rosa, V. Ravì, and R. Rizzo, "Data mining: Classification and prediction," Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, pp. 384–402, 2018, doi: 10.1016/B978-0-12-809633-8.20461-5.
[10]. M. Guarascio, G. Manco, and E. Ritacco, "Deep learning," Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, no. Dl, pp. 634–647, 2018, doi: 10.1016/B978-0-12-809633-8.20352-X.
[11]. M. Fratello, "Decision Trees and Random Forests," pp. 1–10, 2018, doi: 10.1016/B978-0-12-809633-8.20337-3.
[12]. R. G. Brereton, Chemometrics and Statistics: Multivariate Classification Techniques. Elsevier Inc., 2013.
[13]. V. Kotu, and B. Deshpande, "Predictive Analytics and Data Mining - Chapter3: Data Exploration. Predictive Analytics and Data Mining," 2015, doi: 10.1016/B978-0-12-801460-8.00010-0.
[14]. J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification : A measure driven view," Inf. Sci. (Ny)., no. xxxx, 2019, doi: 10.1016/j.ins.2019.06.064.
[15]. P. Galdi and R. Tagliaferri, "Data mining: Accuracy and error measures for classification and prediction," Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, pp. 431–436, 2018, doi: 10.1016/B978-0-12-809633-8.20474-3.
[16]. L. Derczynski, "Complementarity , F-score , and NLP Evaluation," pp. 261–266, 2013. https://doi.org/10.1016/B978-0-12-809633-8.20351-8.
[17]. D. Berrar, "Performance measures for binary classification," Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, no. 1, pp. 546–560, 2018.
[18]. I. Bairamov, "Order Statistics 3rd.pdf," pp. 281–284, 2010.
[19]. R. Tauler and M. Maeder, Multivariate Curve Resolution—Error in Curve Resolution, Second Edition., vol. 2. Elsevier, 2020.
[20]. M. Castelli, L. Vanneschi, and Á. R. Largo, "Supervised learning: Classification," Encycl. Bioinforma. Comput. Biol. ABC Bioinforma., vol. 1–3, no. Ml, pp. 342–349, 2018, doi: 10.1016/B978-0-12-809633-8.20332-4.