Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

An Analysis of Features that Affected Lecturer Scientific Publication Productivity as a Research Output in Indonesia

Ahmad Sanmorino[1], Luis Marnisah[2], Hastha Sunardi[3], Herri Setiawan [4]

[1,3,4]Faculty of Computer Science, Universitas Indo Global Mandiri

[2]Faculty of Economics, Universitas Indo Global Mandiri

e-mail: [1]sanmorino@uigm.ac.id, [2]luismarnisah@uigm.ac.id, [3]hastha_s@uigm.ac.id, [4]herri@uigm.ac.id

Article Info

Page Number: 553 - 571

Publication Issue:

Vol 71 No. 4 (2022)

Article History

Article Received*: 25 March 2022*

Revised*: 30 April 2022*

Accepted*: 15 June 2022*

Publication*: 19 August 2022*

Abstract

This study's objective is to look at the factors that influence academics publication productivity in Indonesia. Each feature is pre-processed using two mechanisms: correlation coefficient (CC) and feature importance calculation (FI). Based on pre-processing, CC and FI scores were obtained for each feature. The three features with the highest CC and FI scores are Number of Published Articles (CC: 0.86, FI: 0.56), Academics H-Score (CC: 0.68, FI: 0.16), and Journal Cluster (CC: 0.65, FI: 0.14). This analysis establishes features that have a significant impact on publication productivity as well as features that have no impact on publication productivity.

Keywords: -Machine learning; feature selection; academics publication productivity; higher education.

Introduction

At the conclusion of a research period, researchers must meet certain objectives known as research results. When it comes to research activities in higher education, the findings demonstrate both the productivity of research in terms of number and quality. [1]. As a result, in order to determine the extent of research progress in higher education, the increase in research productivity, both quantity, and quality, must be measurable [2]. Every country has a mechanism in place to ensure the quantity and quality of research productivity in higher education [3]. The government of Indonesia, through the Ministry of Research and Technology, creates and regulates research mechanisms through a variety of schemes available to lecturers. Each research scheme has different goals, such as

technology implementation, intellectual property rights, and scientific publications. Scientific publications are well-known as a medium for publishing research outcomes and a forum for academics all over the world to share research findings [4]. The number of scientific articles published by lecturers is one indicator of high research productivity in higher education.

The author attempts to analyze the publication productivity of higher education academics in Indonesia through this study. Has the productivity of academic publications in Indonesia met or exceeded the target? With the growing number of scientific publications, a mechanism that can rapidly and massively acquire, store, and analyze publication data is required [5]. The data mining approach is one of the mechanisms used to analyze and benefit from big data from the number of publication [6]. Data mining is an analytical process that uses machine learning, statistics, and databases to find patterns in large data sets. Machine learning is also one of the misnomer approaches [7], because what is actually mined is not the data, but the patterns and knowledge of the data set. Both machine learning and data mining can be used to discover knowledge in a data set [8]. Machine learning has lower complexity than methods that do not require a priori hypotheses, such as data fishing.

Similarly, in higher education, machine learning approaches are the best solution for data analysis of large research publications. A researcher can discover significant variables in publication productivity using the machine learning approach. These variables were used as a framework for developing a mechanism to increase the productivity of scientific publications. A model or framework is used to describe the mechanism for increasing publication productivity. Based on the ease of use and benefits of the machine learning approach, the author employs it to examine the publication productivity of higher education academics in Indonesia. The analysis begins by defining the characteristics that are thought to have an effect on the productivity of publications. The following step is to collect data based on the defined features. The author uses feature selection as a mechanism to select features in order to demonstrate whether the selected features affect the productivity of academic publications. Several machine learning algorithms were used to test the feature selection results. The confusion matrix, accuracy comparison, precision-recall, and AUC score comparison are used to evaluate and analyze the output of modeling. The findings of this machine learning-based feature analysis show which characteristics have the greatest influence on the productivity of academic publications in higher education. Based on this knowledge, the next step is to create a model that incorporates the most effective features in order to increase the

publication productivity of higher education academics in Indonesia. However, the author only discusses the feature analysis in this study; model development will be continued in future studies.

Several related studies on lecturer publication performance are presented in this subchapter. Ramli et al. [9] use a machine learning approach to analyze research performance in higher education. Scientific Papers, Conference, Number of Citations, Age, Gender, Marital Status, Educational Qualifications, Experience, Position, Division were the features used in this study. The researcher employs Logistic Regression, Decision Tree, Artificial Neural Network, and Support Vector Machine classifiers for data modeling. Confusion Matrix, ROC Curve, and Over fitting were used to evaluate the classification results. The Logistic Regression algorithm (Enter Model) achieves an accuracy score of 80.31 percent when tested for classification performance. 83.40 percent for the Decision Tree (Entropy Model). Artificial Neural Network has an accuracy of 82.24 percent, while Support Vector Machine (Linear Kernel) has an accuracy of 80.31 percent.

Wichian et al. [10] looked into the variables that affect the productivity of research in public universities in another related study. Thinking, Research Mind, Volition-Control, International Meeting, The features employed in this study include age, academic level, institutional policy, library expenditure, and resources. Other features include research methodologies, funding of research, and management of research, communication skill, collaboration, and research group. The model is supported by actual data; the Chi-Square value is 80.007. The degree of association between characteristics is determined using the Chi-Square function [11]. The authors evaluate the elements that affect research productivity using BPNN [12]. The accuracy score for the Back Propagation NN classifier analysis was 90.72 percent. This high score demonstrates that the features employed are highly relevant to research productivity.

Henry et al. [13] used five indicators to determine research performance in his study. Because of the large size of the higher education population, primary data were gathered through questionnaires and stratified random sampling. Age cohort, educational qualification, cluster, and lecturer track were discovered to be significant factors in determining academic staff research performance. Awards, job policies, monthly income, research leadership, and research supervisors are all factors that influence research performance. The author employs Logistic Regression to assess academic staff research performance at the higher education. Chi-Square and Nagelkerke R Square were used to evaluate the model's variables. According to Nagelkerke's R Square, the logistic model explains 46 percent of the variation in the outcome variable. The classification evaluation yielded an accuracy score of 78.2

percent.The author developed a mechanism for lecturer productivity in higher institutions in the previous study [14, 15]. The framework has four output variables and nine input variables. Competition, Teamwork, Network, Points, Goals, Inventory, Teammate, Score, and Leveling Up are the nine independent variables that were used, while Sharing Motivation, Competence, Eager Motivation, and Research Publication improvement are the dependent variables. In addition, the Regression between variables indicates whether a variable has good effect. 0.499 is the variable coefficient value, for example, demonstrates that the competition improves academics' learning sharing attitude. This demonstrates that each variable has a strong influence on the others, particularly the dependent variable, Scientific Publication Improvement.

Material and Methods

This study makes use of data collected by the Republic of Indonesia's Ministry of Technology and Research via the Sciences and Technologies Index (SINTA) database. SINTA was introduced in 2017 and has since been actively utilized by lecturers. SINTA gives Indonesians access to citations and scientific knowledge. SINTA is referred to as an information system used to evaluate the performance of researchers, lecturers, and scientific publications in Indonesia on its official website. Furthermore, this platform is a user-friendly publication management system. Another reason, as an online platform, houses lecturer's publication data from whole country. The platform includes a scale for evaluating Indonesian journals and conferences. Before gathering data, the author defines the features that are thought to have an impact on academic publication productivity. Table 1 displays the defined features.

Table 1: The Candidate Features

Features	Description
Number of Published Articles (NP)	Number of the published articles (Low, medium, many)
Journal Cluster (JC)	Journal cluster where the article published (C1 – C4)
Academics H-Score (AH)	Academics H-Index score (High, medium, low)
Academics Gender (AG)	Academics gender (Men, women)
Academics Experience (AE)	Academics research experience (New, expert)
Academics Nationality (AN)	Academics nationality background (Indonesia, Non-Indonesia)
Research Facilities (RF)	Research facilities in higher education (Adequate, Not adequate)
Higher Education Competency (HE)	Type of higher education (Private, public)

Research Level (RL)	Research level obtained by academics (Beginner, basic, advanced/applied)
Research Collaboration (RC)	Number of academics in collaboration (Little, medium, many)
Research Subject (RS)	Academics research subject (Engineering, social, computer, medical)
Position in the Research Team (PR)	Academics position in research team (Leader, member)
Publication Productivity (TAR)	Target to achieve

There are thirteen candidate features that have been predefined. These thirteen features serve as a guide in gathering the data for this study. The author's initial hypothesis is that twelve defined features have a significant effect on Publication Productivity. This hypothesis will be validated through feature selection and machine learning-based analysis. For a more in-depth explanation, then the research design used in this study will be presented. The proposed research design for machine learning-based analysis for features that affected academics publication productivity is shown in Figure 1.

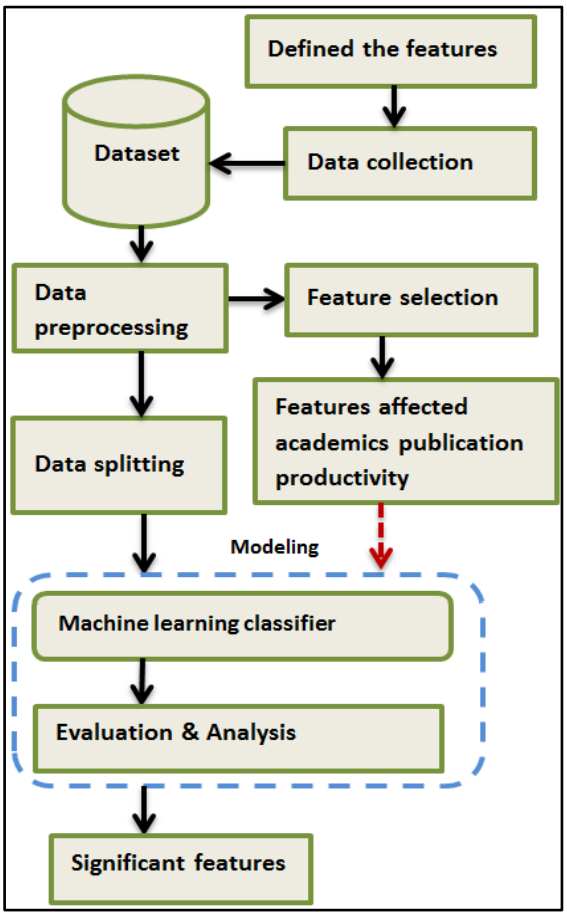

Figure 1: Research Methodology

Preprocessing is the stage that follows data collection. Data Preprocessing is the process of sorting and changing collected data as input for the modeling stage. Table 2 shows the header of the dataset that has passed preprocessing. This header displays the top five records from the entire dataset.

Table 2: Dataset Header

Published Article (NP)	Journal Cluster (JC)	H-Score (AH)	..	Publication Productivity
8	9	7	..	6
8	9	8	..	6
6	6	6	..	5
7	8	7	..	6
8	8	7	..	6

The measurement level for dependent features is binary. Six indicates that academics meet the Publication Productivity target, five indicates that academics do not meet the Publication Productivity target. These two values can also be interpreted as high and low publication productivity, respectively. Preprocessed data is also used to select features that affect academics publication productivity. The feature selection results are used as a guide in analyzing features that influence publication productivity [16,17]. This means that not all preprocessed features or data are used for modeling or analysis.

In feature selection, two mechanisms are used. The first mechanism, Feature Importance, is used to calculate a feature's relevance score. The higher a feature's score, the more important it is to the target feature. In other words, Feature Importance is used to determine the level of information (gain) contained in a feature or variable [18]. Equation (1) and (2) are used to calculate the feature importance score.

$$Entropy(S) = \sum_{i=1}^{C} p_i \, log_2(p_i) \quad (1)$$

Where c denotes the number of distinct class labels and pi denotes the proportion of rows with output label i. After obtaining the entropy, it is entered into the equation (2).

$$Gain(S, A) = Entropy(S) - \sum_{veValues\ (A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

(2)

Where S denotes the set of instances, A denotes the attribute, Sv denotes the subset of S with A = v, and Values (A) denotes the set of all possible values of A. The Correlation Coefficient is the second feature selection mechanism. The correlation coefficient represents the relationship between independent features and other independent or dependent features. The correlation score can be advantageous if the open feature increases influences an improvement in the dependent feature or vice versa. The correlation coefficient score is calculated using the following ccorrelation coefficient equation [19].

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

(3)

Where n denotes the number of pairs of scores, xy denotes the number of products of the paired scores, x denotes the number of scores x, y denotes the number of scores y, x2 denotes the number of scores x squared, y2 denotes the number of scores y squared. The data is divided into two groups as input for the machine learning classifier: the training set and the testing set [20, 21]. Support Vector Machine, Multilayer Perceptron, Random Forest [22], and Naive Bayes are the four machine learning classifiers used in the analysis [23]. To evaluate the modelling results, the confusion matrix [24], accuracy score comparison [25], precision-sensitivity score comparison, and the comparison of area under the curves were used. Based on the results of the evaluation, it will be determined whether or not the features have a significant impact on academic publication productivity. In the end, the author also compares this study's to other same studies, to find out whether the conclusions of this research are significant or not.

Results and Discussion

Following preprocessing, three candidate features are not used: Academic Experience, Research Facilities, and Position in Research Team. The author is having difficulty obtaining accurate data for these three features. There are nine independent features and one dependent feature for the next stage. The discussion begins with the selection of features. The Correlation Coefficient scores are

calculated using a library in Python programming. The correlation coefficient is visualized by heat maps. Heat maps show the relationship between independent features to other independent features or to dependent features. Figure 2 shows the heat maps of the Correlation Coefficient for all of the features.

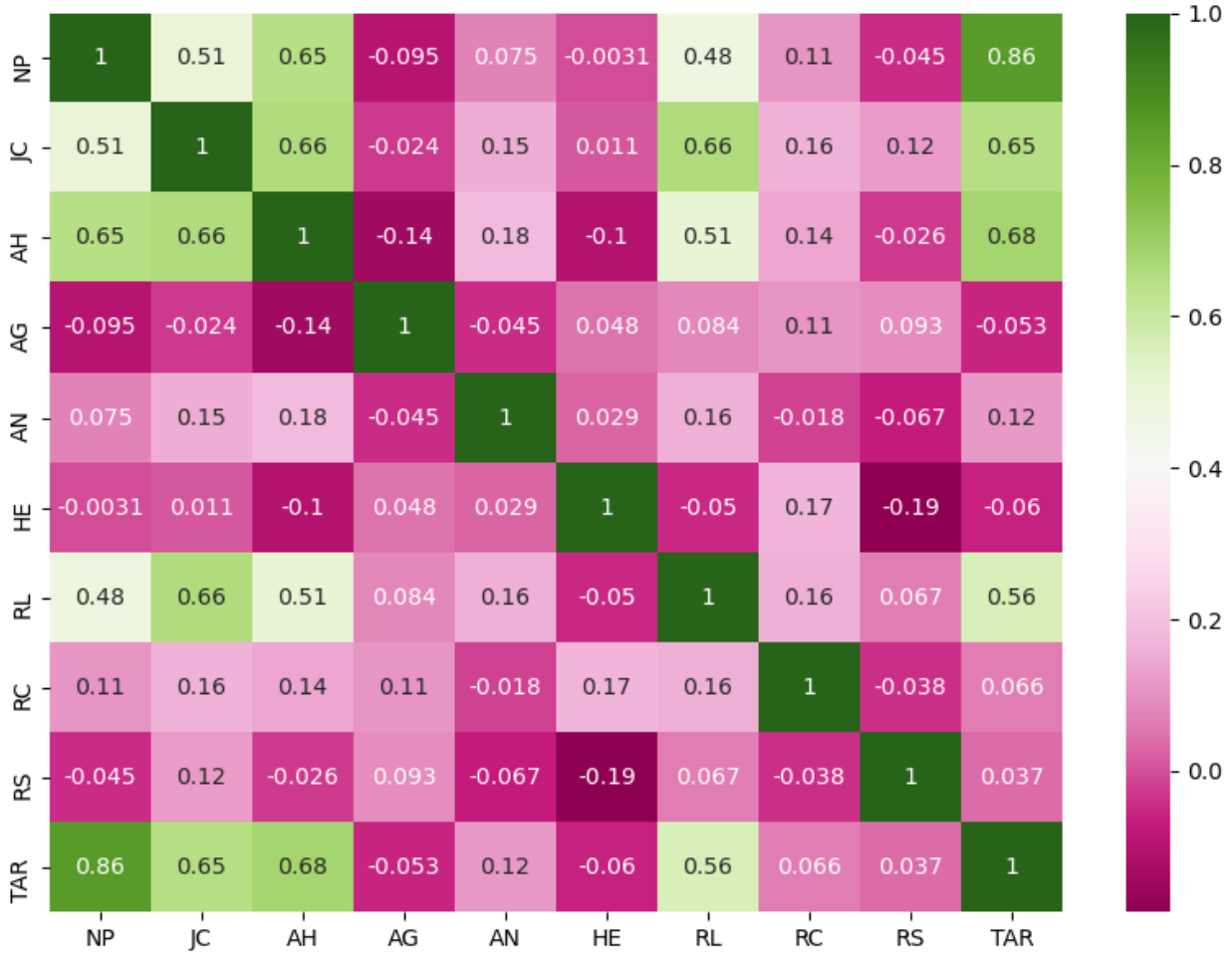

Figure2: The Correlation Coefficient heat maps

A correlation score of one or close to one indicates the best correlation. A score of 0 or negative indicates that there is little or no correlation between the features. The Number of Published Articles (NP) has the highest correlation score on Publication Productivity, at 0.86, according to heat maps. Meanwhile, with a score of -0.06, Higher Education Competency (HE) is at the bottom (Table 3). After being tested, some features that were previously thought to have a high correlation score turned out to have no correlation (zero correlation or even negative correlation) on Publication Productivity.

Table3: Correlation Coefficients contribute to Publication Productivity

Features	Correlation Coefficient Score
Number of Published Articles (NP)	0.86
Academics H-Score (AH)	0.68
Journal Cluster (JC)	0.65
Research Level (RL)	0.56
Academics Nationality (AN)	0.12
Research Collaboration (RC)	0.066
Research Subject (RS)	0.037
Academics Gender (AG)	-0.053
Higher Education Competency (HE)	-0.06

The selection process in the second mechanism, feature importance, begins with the calculation of entropy. The entropy score is entered into the Feature Importance (FI) equation. Table 4 shows the FI score.

Table 4: The Features Importance score

Features	Feature Importance Score
Number of Articles (NP)	0.56
Academics H-Score (AH)	0.16
Journal Cluster (JC)	0.14
Research Level (RL)	0.09
Research Subject (RS)	0.02
Academics Gender (AG)	0.01
Higher Education Competency (HE)	0.009
Academics Nationality (AN)	0.0006
Research Collaboration (RC)	0.0001

The highest feature importance score was obtained by the Number of Published Articles (NP), which was 0.56. The higher the score, the more important a feature is in relation to the target feature. Academic Nationality (AN) and Research Collaboration (RC) are ranked lowest, indicating that these two characteristics are less important to Publication Productivity. Following the calculation of

the correlation score and the Feature Importance of each feature, an analysis involving several machine learning classifiers is performed. First, the modeling results are assessed using a confusion matrix that includes True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Table 5 shows the evaluation results.

Table 5: The evaluation using confusion matrix

Classifiers	TP	FP	TN	FN
Support Vector Machine	36.66%	0.00%	56.67%	6.67%
Multilayer Perceptron	33.33%	3.33%	63.33%	0.00%
Random Forest	30.00%	0.00%	63.33%	6.67%
Naive Bayes	30.00%	0.00%	33.33%	36.67%

56.67% of correctly recognized academics did not fulfill the publishing productivity objective, according to the evaluation utilizing the Confusion-Matrix for the Support Vector Machine. 6.67 percent of academics were mistakenly classified as not producing enough publications. 37.66% of academics accurately identified themselves as having achieved the publication productivity goal. Academics were mistakenly classified as not meeting the target for publication productivity in 0% of cases. According to the analysis of the Multilayer Perceptron utilizing the Confusion-Matrix, 63.33 percent of correctly identified academics fell short of the target for publishing productivity. 0% of academics were mistakenly classified as falling short of the goal for publication productivity. The professors who accurately identified themselves as having achieved the publishing productivity target made up 33.33 percent. Academics that were mistakenly classified as meeting the target for publication productivity comprised 3.33 percent of the total.

According to the Random Forest evaluation utilizing the Confusion-Matrix, 63.33 percent of correctly identified academics fell short of the target for publishing production. 6.67 percent of academics were mistakenly classified as not producing enough publications. Only 30% of academics accurately described themselves as having achieved the publication productivity goal. Academics were mistakenly classified as not meeting the target for publication productivity in 0% of cases. According to the analysis utilizing the Confusion-Matrix for the Nave Bayes, 33.33 percent of correctly recognized academics fell short of the goal for publishing productivity. The objective for publication productivity was wrongly set at 36.67 percent of academics. Only 30% of academics

accurately described themselves as having achieved the publication productivity goal. Academics were mistakenly classified as not meeting the target for publication productivity in 0% of cases.

The accuracy score, precision-sensitivity, and f1-score of each classifier are shown in Table 6 – Table 9. Table 6 shows the classification report of the Support Vector Machine algorithm.

Table 6: Support Vector Machine classification report

	fidelity	Sensitivity	F-measure
Zero(0)	85%	100%	92%
One(1)	100%	89%	94%
Acc			93%
MA	92%	95%	93%
WA	94%	93%	93%

85 percent of the academics who were predicted by the Support Vector Machine (SVM) to fail did not reach the publishing productivity goal. Academics who are projected by SVM to reach the aim for publication productivity are produced at a rate of 100%. Compared to all academics who do not meet the publishing productivity target, SVM produces 100% of academics who are expected to not meet the target. When compared to all academics that actually fulfill the publishing productivity objective, SVM generates 89 percent of those who are projected to do so. For academics that fall short of the goal of 92 percent publishing productivity, SVM computes a comparison of the average precision and sensitivity. Academics that meet the 94 percent publishing productivity threshold have their average precision and sensitivity computed by SVM (f1-Score). SVM generated 93% of academics who were accurately predicted to fulfill the publishing productivity target, although not all academics achieved the target (accuracy). The Multilayer Perceptron algorithm's classification report is displayed in Table 7.

Table 7: Multilayer Perceptron classification report

	fidelity	Sensitivity	F-measure
Zero(0)	100%	91%	95%
One(1)	95%	100%	97%
Acc			97%
MA	97%	95%	96%

WA	97%	97%	97%

One hundred percent of the academics predicted by the Multilayer Perceptron (MP) to fail did not reach the publishing productivity objective. Academics who are expected to reach the target for publication productivity come from MP in 95% of cases. Comparing MP to all academics who do not meet the publishing productivity target, MP produces 91% of academics who are anticipated to not meet the target. When compared to all academics who actually fulfill the publishing productivity target, MP generates 100% of those who are anticipated to do so. For academics that fall short of the goal of a 95 percent publishing productivity, MP computes a comparison of the average precision and sensitivity. Academics that meet the 97 percent publishing output objective have their average precision and sensitivity compared by MP (f1-Score). While not all academics produced by MP met the publication productivity target, 97 percent of those who were correctly predicted to do so did so (accuracy). The classification report of the Random Forest algorithm is displayed in Table 8.

Table 8: Random Forest classification report

	fidelity	Sensitivity	F-measure
Zero(0)	82%	100%	90%
One(1)	100%	90%	95%
Acc			93%
MA	91%	95%	93%
WA	95%	93%	94%

Eighty-two percent of the academics who were predicted to fail by the Random Forest (RF) did not reach the publication productivity objective. Academics who are anticipated to reach the publishing productivity target are produced by RF in full force. Compared to all academics who do not meet the publishing productivity target, RF produces 100% of academics who are expected not to meet the target for publication productivity. When compared to all academics who actually fulfill the publishing productivity target, RF generates 90% of those who are anticipated to do so. For academics who fall short of the goal of 90 percent publishing productivity, RF computes a comparison of the average precision and sensitivity. Academics who meet the 95 percent publishing production threshold have an average precision and sensitivity calculated by RF (f-measure). Ninety-three percent of academics produced by RF were accurately expected to fulfill the publishing

productivity target, although not all academics did (accuracy). The Nave Bayes algorithm's classification report is displayed in Table 9.

Table 9: Naïve Bayes classification report

	fidelity	Sensitivity	F-measure
Zero(0)	45%	100%	62%
One(1)	100%	48%	65%
Acc			63%
MA	72%	74%	63%
WA	84%	63%	64%

Of all academics predicted to fail, the Nave Bayes (NB) yielded 45% who did not fulfill the publishing productivity target. Academics anticipated to meet the target for publishing productivity are entirely produced by NB. Compared to all academics who do not meet the publishing productivity target, NB generates 100% of academics who are anticipated not to meet the target for publication productivity. When compared to all academics that actually fulfill the publishing productivity objective, NB produces 48% of those who are anticipated to do so. For academics that fall short of the goal of 62 percent publishing productivity, NB computes a comparison of the average precision and sensitivity. Academics who meet the 65 percent publishing output target have their average precision and sensitivity calculated by NB (f1-Score). Although NB produced 63% of academics who were forecasted to meet the publication productivity target correctly, they did not all meet the publication productivity target (accuracy).

The area under the curve (AUC) is calculated using the Receiver Operating Characteristic curve (ROC curve) in the following analysis, the better the outcomes of machine learning modeling, the larger the area under the curve. Figure 3 shows the area under the curve for each classifier.

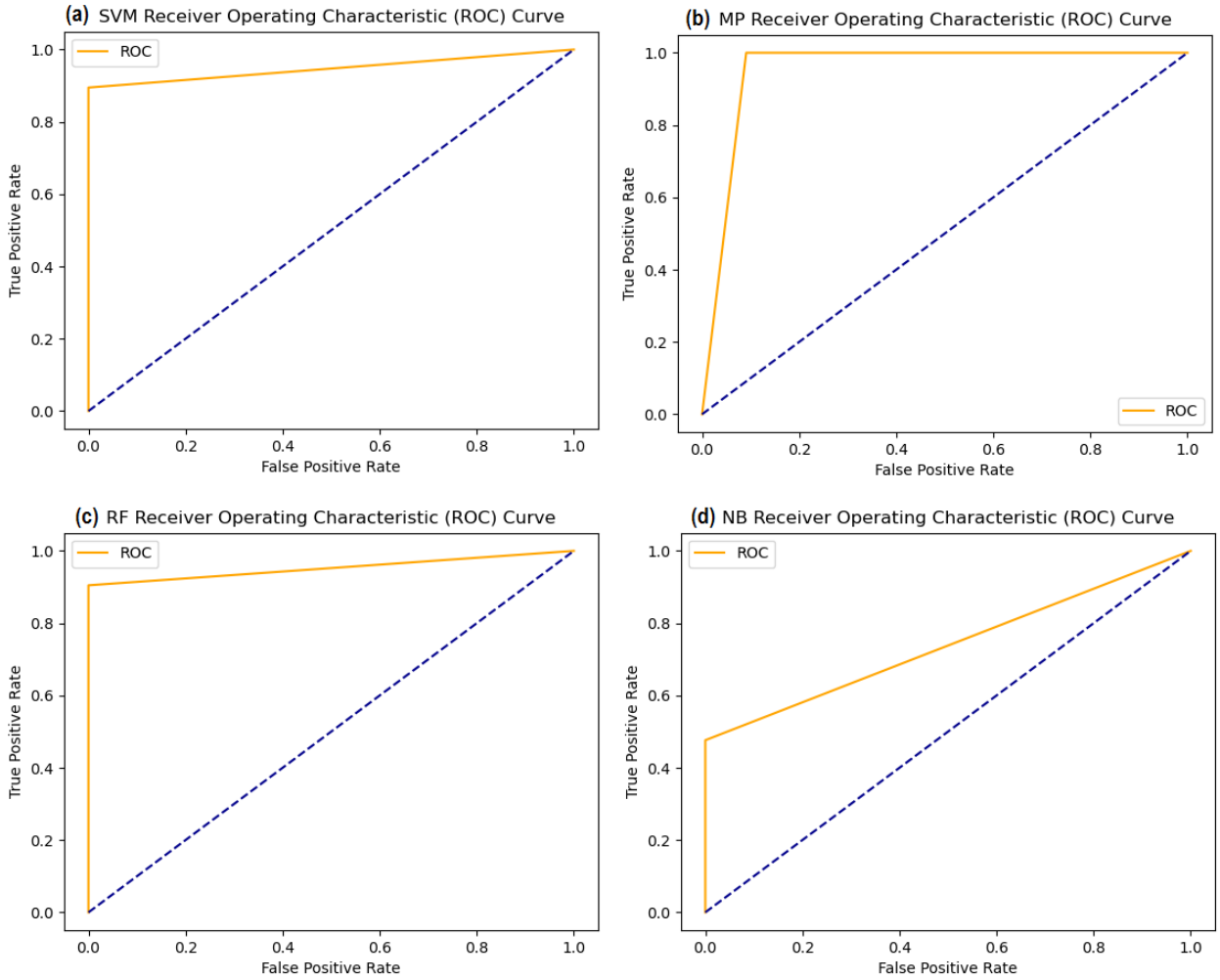

Figure 3: (a) SVM ROC Curve; (b) MP ROC Curve; (c) RF ROC Curve; (d) NB ROC Curve

SVM has an AUC of 95 percent, MP has an AUC of 98 percent, RF has an AUC of 95 percent, and NB has an AUC of 74 percent. These results show that the Multilayer Perceptron analysis achieves the highest AUC score when compared to other classifiers, as well as that the analyzed features have a significant impact on Publication Productivity. Each classifier's precision and misclassification rate are compared in Table 10.

Each classifier has an accuracy score greater than 70 percent. This demonstrates that the tested features are highly relevant and have a significant impact on academic publication productivity.

Table 10: Accuracy score comparison

Classifier	Accuracy	Misclassification Rate
SV Machine	93%	7%

Multilayer PP	97%	3%
Random F	93%	7%
Naïve Bys	63%	37%

The goal of this evaluation is to see if the analysis using the machine learning classifier can exceed 70 percent score, rather than find the highest score. This evaluation also demonstrated that the analyzed features had a positive impact and could be used for a variety of purposes in the future.Table 11, the author compares some evaluation results from other related studies.The quantity of datasets utilized, the combination of methods, and the variables employed can all alter the outcome of an analysis, but this study has shown that each element has a substantial impact on the productivity of academic papers. The authors cannot claim that the results of this test are superior to those of other similar studies. The same scenario must be utilized to show that because it can alter the test findings. This means that the location of data collection, the number of datasets, the method for picking variables, and the number of variables must all be the same.

Table 11: The accuracy scores comparison, toward other related studies

Author	Features Name	Feature Selection Mechanism	Classifier	Accuracy
Ramli et al. [9]	Age, Gender, Marital Status, Education, Work Experience, Position, Division, Citation, and Target are all factors in an article (Status of Research Performance)	Not Mentioned	DT ANN Logistic Regression SV Machine	83.40% 82.24% 80.31% 80.47%
Henry et at. [13]	Success, Workplace Policy, Monthly Income, Age Cohort, Highest Qualification, Cluster, Lecturer Track, and Research Leadership	Chi-Square, Nagelkerke R Square, Lemeshow test	Logistic Regression	78.2%

Wichian et al. [10]	Age, academic position, thinking, research mind, volition-control, international meeting, institutional policy, library expenditure, networking and teamwork, research management, Techniques for conducting research, research funding	Chi-Square, R-Square, Cronbach Alpha	BPNN	90.72%
Zakree et al. [26]	Grant amount, department, administrative position, number of PhD students, faculty, keynote speaker invitation, article (index), age, designation, number of research grants, gender, performance score Working status and marital status	Spearman Rho Correlation	PART J-48 C4.5 Decision Tree	75.00% 75.30% 70.20% 70.30%
Sanmorino et al. (this study)	Journal Cluster, Research Level, Research Subject, Academic Gender, Higher Education Competency, Academic Nationality, Research Collaboration, and Publication Productivity are some indicators of the number of articles published (Target)	Correlation Coefficient, Feature Importance	SV Machine Multilayer PP Random F Naïve Bys	93.00% 97.00% 93.00% 63.00%

Conclusion

The effect of each feature on publication productivity is known based on feature selection, evaluation, and analysis using machine learning. The scores obtained for each feature using correlation coefficient (CC) and feature importance (FI) are: Number of Published Articles (CC: 0.86, FI: 0.56), Academics H-Score (CC: 0.68, FI: FI: 0.16), Journal Cluster (CC: 0.65, FI: 0.14), Research Level (CC: 0.56, FI: 0.09), Research Subject (CC: 0.037, FI: 0.02), Academics Gender (CC: -0.053, FI: 0.01), Higher Education Competency (CC: -0.06, FI: 0.009), Academics Nationality (CC: 0.12, FI: 0.0006), Research Collaboration (CC: 0.066, FI: 0.0001). In addition, the authors compared the accuracy scores, with the exception of Naive Bayes, with each classifier scoring above 70 percent. This analysis establishes features that have a significant impact on publication productivity as well as features that have no impact on publication productivity. In the future, features that have a significant impact on publication productivity will be used as the construct for a model to increase publication productivity in higher education.

Acknowledgment

This research is fully funded by the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for PTUPT Grant 2022, which made this research endeavor possible.

References

[1] Fauzi, M.A., Nya-Ling, C.T., Thursamy, R., Ojo, A.O. (2019). Knowledge sharing: Role of academics towards research productivity in higher learning institution. VINE J. Inf. Knowl. Manag. Syst. 49, 136–159.

[2] Abramo, G., D'Angelo, C.A. (2014). How do you define and measure research productivity?. Scientometrics 101, 1129–1144.

[3] Wills, D., Ridley, G., Mitev, H. (2013). Research productivity of accounting academics in changing and challenging times. J. Account. Organ. Chang. 9, 4–25.

[4] Dhillon, S.K., Ibrahim, R., Selamat, A. (2015). Factors associated with scholarly publication productivity among academic staff: Case of a Malaysian public university. Technol. Soc. 42, 160–166.

[5] Addo-Tenkorang, R., Helo, P.T. (2016). Big data applications in operations/supply-chain management: A literature review. Comput. Ind. Eng. 101, 528–543.

[6] Aggarwal, C.C. (2015). Data Mining: The Textbook. Springer Publishing Company, Incorporated.

[7] Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., Cattrysse, D. (2015). Literature Review of Data Mining Applications in Academic Libraries. J. Acad. Librariansh. 41, 499–510.

[8] Alzubaidi, A., Tepper, J., Lotfi, A. (2020). A novel deep mining model for effective knowledge discovery from omics data. Artif. Intell. Med. 104, 101821.

[9] Ramli, N.A., Nor, N.H.M., Khairi, S.S.M. (2019). Prediction of research performance by academicians in local university using data mining approach, 04, 0021.

[10] Na Wichian, S., Wongwanich, S., Bowarnkitiwong, S. (2009). Factors affecting research productivity of faculty members in government universities: LISREL and Neural Network analyses. Kasetsart J. - Soc. Sci. 30, 67–78.

[11] Molugaram, K., Rao, G.S. (2017). Chapter 9 - Chi-Square Distribution, in: Molugaram, K., Rao, G.S.B.T.-S.T. for T.E. (Eds.), . Butterworth-Heinemann, pp. 383–413.

[12] Zupan, J.B.T.-D.H. in S. and T. (2003). Basics of artificial neural networks, in: Nature-Insprired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks. Elsevier, pp. 199–229.

[13] Henry, C., Md Ghani, N.A., Hamid, U.M.A., Bakar, A.N. (2020). Factors contributing towards research productivity in higher education. Int. J. Eval. Res. Educ. 9, 203–211.

[14] Sanmorino, A., Ermatita, Samsuryadi. (2019). 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. 8, 554–559.

[15] Sanmorino, A., Marnisah, L., Sunardi, H. (2021). A gamification framework for research productivity enhancement on the higher education institution. Int. J. Eval. Res. Educ. 10, 706–713.

[16] Setiawan, H., Husnawati, & Tasmi. (2021). Assessment System of Local Government Projects Prototype in Indonesia. International Journal of Advanced Computer Science and Applications, 12(12), 425–432.

[17] Tasmi, Setiawan, H., Stiawan, D., Husnawati, & Valiata, S. A. (2019). Determining attributes of encrypted data traffic using feature selection method. International Journal of Engineering and Advanced Technology, 9(1), 3500–3504. https://doi.org/10.35940/ijeat.A2674.109119

[18] Kareva, I., Karev, G. (2020). Replicator dynamics and the principle of minimal information gain. pp. 129–154.

[19] Franzese, M., Iuliano, A. (2018). Correlation analysis. Encycl. Bioinforma. Comput. Biol. ABC Bioinforma. 1–3, 706–721.

[20] Urso, A., Fiannaca, A., La Rosa, M., Ravì, V., Rizzo, R. (2018). Data Mining: Classification and Prediction.

[21] Guarascio, M., Manco, G., Ritacco, E. (2018). Deep learning. Encycl. Bioinforma. Comput. Biol. ABC Bioinforma. 1–3, 634–647.

[22] Fratello, M., Tagliaferri, R. (2018). Decision Trees and Random Forests.

[23] Kotu, V., Deshpande, B. (2015). Chapter 3 - Data Exploration, in: Kotu, V., Deshpande, B.B.T.-P.A. and D.M. (Eds.). Morgan Kaufmann, Boston, pp. 37–61.

[24] Xu, J., Zhang, Y., Miao, D. (2019). Three-way confusion matrix for classification : A measure driven view. Inf. Sci. (Ny).

[25] Galdi, P., Tagliaferri, R. (2018). Data Mining: Accuracy and Error Measures for Classification and Prediction.

[26] Zakree, M., Nazri, A., Ghani, R.A., Abdullah, S., Ayu, M., Samsiah, R.N. (2019). Predicting Academician Publication Performance using Decision Tree. 180–185.