Impact of H-Index Toward Citations Using Linear Regression on Science and Technology Index

Dihin Muriyatmoko*[1], Lalu Ganda Rady Putra[2]
[1,2] Universitas Darussalam Gontor
dihin@unida.gontor.ac.id[*]

Abstract

Number of Journals in Indonesia is quite a lot and various disciplines. Until March 15, 2018, registered 50,889 online and print ISSN by Indonesian Institute of Science (LIPI). The government through Ministry of Research, Technology, and Higher Education of Republic Indonesia (Kemenristekdikti) set regulated on journal index, that is Science and Technology Index (Sinta) assigned to rank quality content and management divided by six categories called S1 to S6 which of the data is taken from Google Scholar and Scopus. This research applies S1 that these journals is accredited "A" by Kemenristekdikti and or index by Scopus. That's data is shown ranking by sorted based on h-index and citations. S1 shown that journal which has highest h-index uncertain have highest citations too, even some have zeroes. That's data on S1 become strange and awkward when compared with S2 to S6 because some value of h-index and citations S1 is lower than S2 to S6. This research focus to quantify how strong correlation and impact h-index toward citations using linear regression. The test result show that value of Multiple R = 0.78 is included high correlation, value of R Square = 0.61 indicates the impact of h-index toward citations achieve 61% and the rest 39% affected by others factor.

Keywords: *Linear Regression; Citation; H-Index; Sinta; Science and Technology Index*

1. Introduction

Writing and publishing scientific paper now is the obligation requirement for lecturer in Indonesia which one of three to supporting conditions of college Tridharma. Refer to Law of the Republic Indonesia, Number 12 Year 2012 on Higher Education at chapter 1 article 1 verse 1, Tridharma is the obligation of college to organize education, research, and community service. Accordingly, Indonesian lecturer is a professional educator and scientists with primary task to transforming, developing, and disseminating science and technology through education, research and community service. Therefore a lecturer must carry out the publication of articles in the journal [1]. The number of national lecturer currently reach 282.486 people [2], while the number of national scientific journals is also quite a lot of disciplines, registered on International Standards of Serial Number (ISSN) Indonesian Institute of Sciences (LIPI: Lembaga Ilmu Pengetahuan Indonesia) reached 50,889 (2) both ISSN online and print [3]. To manage the quality of content and management of scientific journals, the government through Ministry of Research, Technology, and Higher Education of Republic Indonesia (Kemenristekdikti) has published categorization into six types namely Sinta 1 (S1) to Sinta 6 (S6), which is the main data for the calculation of articles and citations derived from Scopus and Google Scholar [4] are ranked based on the value of highest h-index to lowest.

H-index is one alternative to the journal impact factor to qualify journals [5], besides it the h-index is an indicator widely used to assess the quality of researchers and organizations [6]. This research takes data from S1 which the highest ranking with criteria of accredited scientific journals is "A" from Kemenristekdikti and or indexed in Scopus International Reputable Indexing. The data listed indicate the h-index and the number of different citations. In the data, high h-index does not necessarily have a high number of cites, there is even a h-index and a zero citation. This raises a bit of an awkwardness for a science journal S1 that should be the most desirable but has h-index and fewer citation counts than the underlying category scientific journals. This is the background of researchers to deeper information that aims to measure the extent of the impact or correlation of h-index toward the number of citations. This research uses linear regression method to test the causal correlation between two variables that is h-index toward citations.

Here are some related research h-index, citations and linear regression; the first measuring the h-index was found to be robust to changes in citations up to approximately the 25th percentile of the citation distribution, inflating its value afterwards [5], second association of h-index of editorial boards member and impact factor radiology journals show that editorial boards member of radiology journals with impact factor the high has h-index lower than journal impact factor [7], third from forestry journals ranking based on the indexing data from ISI Web of Science (WOS) and Google Scholar which states that h-index have a correlation or high relationship with Journal Impact Factor and ranking of 180 forestry journal presented using this indexation [8], fourth measurement results h-index from 455 journal in business and management indexed ISI Web Of Science (WOS) and Association of Business School's peer review journal ranking list show that h-index is preferable to impact factor for a variety of reasons, mainly selective coverage IF and disadvantages journal if many papers are published [9]. Fifth interdisciplinary high paper show that top 1% most high cited papers interdisciplinarity than articles in other citation [10]. Sixth citation of articles on software engineering study found that number of articles published per year has currently 6000–7000 articles are published every year and grown tremendously, so currently the hot topics in software engineering are: 1. web services, 2. mobile and cloud computing, 3. industrial (case) studies, 4. source code and 5. test generation [11]. Seventh literature review on citation impact indicator show that bibliometricians should expand their perspective on citation analysis to take advantage opportunities by new data sources [12]. Eighth estimation of onshore wind resources in the coast, Brazil using linear regression show well performance in terms of high level of agreement data series, predicted data and associate correlation coefficients [13]. Ninth stock market prediction using linear regression to forecasting behavior of TCS data set, evident that proposed method is optimum to compare the other regression technique [14]. Tenth estimate budget house construction cost using linear regression model is fast, easy, and accurate, so decision for investment advisability can be completed very fast [15]. Eleventh temperature prediction using multiple linear regression shown that the value of prediction is accurate and measures should be taken to avoid the temperature will terrace to an unlivable level [16].

2. Research Method

This research used linear regression method that can generate prediction of value between predictor and response variable. Beside it, linear regression method can use to explanatory the level and characteristics of relationship between the dependent and independent variable [17], the method step is explained on Figure 1.
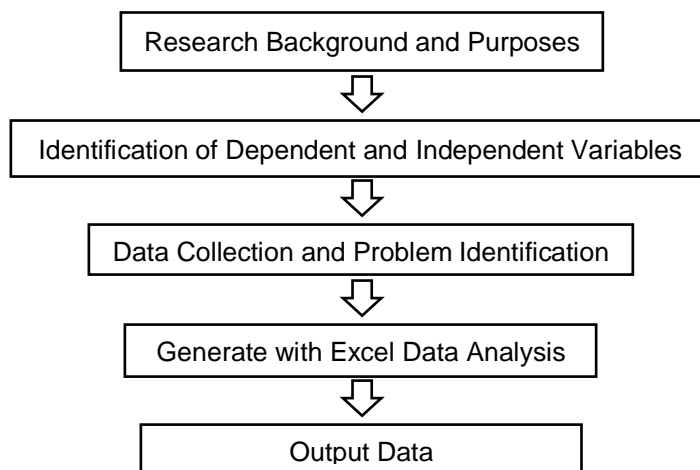

Figure 1. Step of Linear Regression Method

2.1 Research Background and Purpose

The basic idea behind this research was an intention to the researcher as editor of Fountain of Informatics Journal [18] and his journal include on S3 with zero h-index and zero citation [4]. The Journal data on Sinta displayed based of highest to lower h-index but on another reputable journal ranking such as Scimagojr does not displayed based on h-index [19], so this causes curiosity to researcher. The research purpose focus to quantify how strong correlation between h-index and citation, and to measure impact of h-index toward citation based on S1.

2.2 Identification of Dependent and Independent Variables

Refer to Sinta journal data, the dependent variable (Y) is number of citation and independent variable (X) is h-index.

2.3 Data Collection and Problem Identification

This research take data from Science and Technology Index with S1 journals category, that these journals is accredited "A" by Kemenristekdikti and or index by Scopus accessed date on January, 19, 2018. This dataset shown on Table 1 sort by h-index descending and number of S1 is fourthly (40) journals and dataset from S2 with some journals order by h-index descending. Science and technology index have self-role or private, but not open access, so this research only use the h-index and citations number which exist in his website [4].

Table 1. Journal by Science and Technology Index 1 (S1) sort by Journals h-index [4]

No	Journal Name	H-index	Citations
1	Gadjah Mada International Journal of Business (GamaIJB)	24	2150
2	TELKOMNIKA (Telecommunication Computing Electronics and Control)	24	3877
3	Indonesian Journal of Electrical Engineering and Computer Science	21	5939
4	International Journal of Electrical and Computer Engineering	21	4473
5	Media Peternakan	20	1909
6	Studia Islamika	19	1592
7	HAYATI Journal of Biosciences	17	1469
8	International Journal of Power Electronics and Drive Systems	16	1615
9	Jurnal Pendidikan IPA Indonesia	16	1049
10	Kukila	16	885
11	Bulletin of Chemical Reaction Engineering & Catalysis	15	1041
12	Biodiversitas	14	514
13	International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)	14	1554
14	AGRIVITA, Journal of Agricultural Science (AJAS)	13	1327
15	Journal of the Indonesian Tropical Animal Agriculture	13	775
16	Microbiology Indonesia	13	957
17	Indonesian Journal of Chemistry	12	1026
18	Medical Journal of Indonesia	12	1475
19	Jurnal Manajemen Hutan Tropika	11	666
20	Indonesian Journal of Applied Linguistics	10	414
21	Indonesian Journal of Applied Linguistics	10	414
22	Journal of Engineering and Technological Sciences	10	618
23	Journal of Mathematical and Fundamental Sciences	10	535
24	Indonesian Journal of Geography	9	635
25	Journal of ICT Research and Applications	9	365
26	Al-Jami'ah: Journal of Islamic Studies	8	429
27	Bulletin of Electrical Engineering and Informatics	7	346
28	Journal of Indonesian Islam	7	259
29	Electronic Journal of Graph Theory and Applications	6	132
30	Indonesian Journal of Electrical Engineering and Informatics	5	169
31	Paramita: Historical Studies Journal	5	110
32	IJAIN (International Journal of Advances in Intelligent Informatics)	4	53
33	Indonesian Journal of Islam and Muslim Society	3	50
34	ACTA MEDICA INDONESIANA	0	0
35	BIOTROPIA	0	0
36	Critical Care and Shock	0	0
37	International Journal of Electrical Engineering and Informatics	0	0
38	International Journal of Technology	0	0
39	Jurnal Respirologi Indonesia	0	0
40	TEFLIN Journal	0	0

Refer to Table 2, taken ten (10) journals sort by journal h-index (accessed date on March, 16, 2018), indicate that h-index and citations is higher than Table 1, even though ranking by Kemenristekdikti the S1 is more upper than S2.

Refer to Figure 2 and Figure 3, the data compared that h-index and number of citation on S2 is higher than S1, but actually the case the S1 data are highest quality nationally on content and management. In S1 the highest h-index value is 24 with citation value 2150 and S2 the highest h-index value is 28 with citation value 3118.

Table 2. Journal by Science and Technology Index 2 (S2) Sort by Journals h-index [4]

| No	Journal Name	H-index	Citations
1	Jurnal Manajemen dan Kewirausahaan	28	3118
2	Journal of Indonesian Economy and Business (JIEB)	26	2716
3	k@ta	25	3613
4	Jurnal Akuntansi dan Auditing Indonesia	23	1917
5	Makara Journal of Health Research	21	1794
6	Jurnal Akuntansi dan Keuangan Indonesia	21	1555
7	Makara Journal of Science	21	1531
8	Jurnal Akuntansi dan Keuangan	20	2399
9	Civil Engineering Dimension (Dimensi Teknik Sipil)	19	1771
10	Makara Human Behavior Studies in Asia	17	1094
11	Makara Journal of Technology	17	796
12	Buletin Penelitian Kesehatan	16	1863
13	JURNAL KEUANGAN DAN PERBANKAN	16	1273
14	Agritech	16	1096
15	Jurnal Teknologi dan Industri Pangan	16	1016
16	JURNAL DINAMIKA MANAJEMEN	16	1003
17	Jurnal Iktiologi Indonesia	16	936
18	Bulletin of Monetary Economics and Banking	16	909
19	Buletin Ekonomi Moneter Perbankan	16	909
20	Paediatrica Indonesiana	15	1440
21	Pharmaceutical Sciences and Research (PSR)	15	1152
22	Indonesian Journal of Marine Sciences	15	1115
23	Jurnal Veteriner	15	1084
24	Jurnal Agronomi Indonesia (Indonesian Journal of Agronomy)	15	905
25	Jurnal Siasat Bisnis	15	799
26	Sari Pediatri	14	1809
27	Jurnal Ilmu dan Teknologi Kelautan Tropis	14	783
28	Dental Journal (Majalah Kedokteran Gigi)	13	1060
29	INDONESIAN JOURNAL OF PHARMACY	13	1042
30	Indonesian Journal of Pharmacy	13	1042
31	Jurnal Antropologi Indonesia	13	829
32	Economic Journal of Emerging Markets	13	795
33	JURNAL PSIKOLOGI	13	683
34	Jurnal Gizi dan Pangan	13	642
35	Ilmu Pertanian (Agricultural Science)	13	590
36	Kesmas: National Public Health Journal	12	814
37	Anima Indonesian Psychological Journal	12	688
38	KEMAS Journal: Research Study in Public Health	12	514
39	Jurnal Gizi Klinik Indonesia	12	512
40	Buletin Peternakan	11	1082
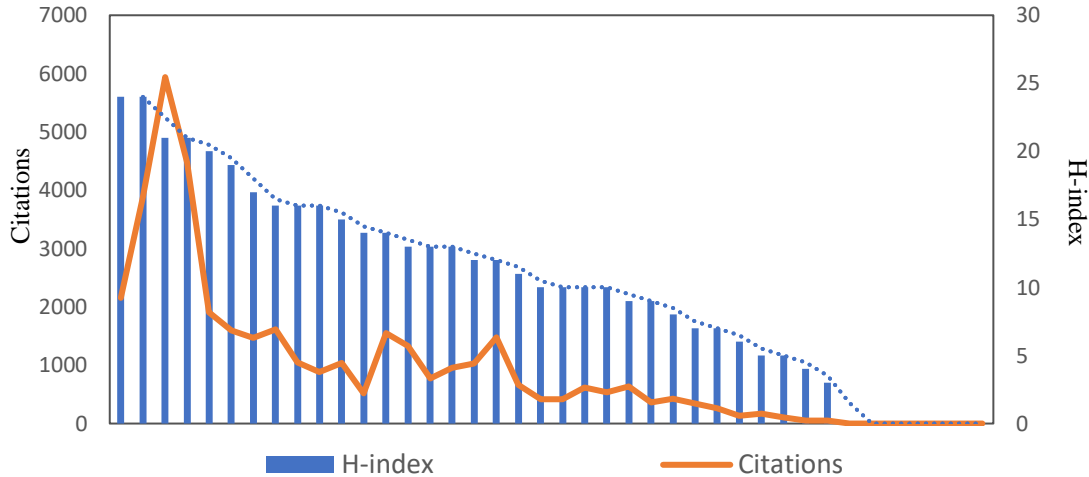
Figure 2. H-index and Citation of 40 Journal on S1

Refer on Figure 2. The S1 shown that the highest h-index not certainly highest number of citation too, even some have zeroes.
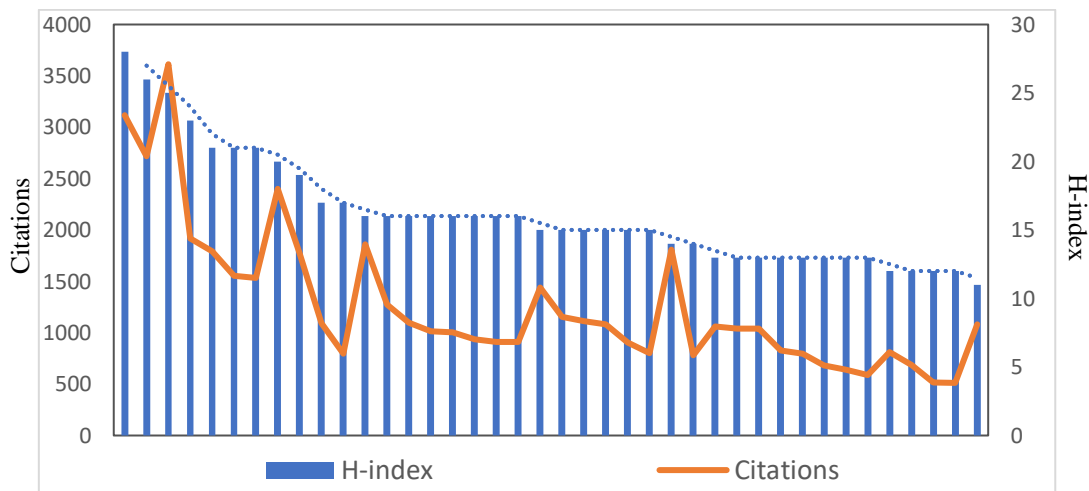

Figure 3. H-index and Citation of 40 Journal on S2

2.4 Generate with Excel Data Analysis

This research focus to measure how strong correlation h-index journal toward number of citations and how many percentages factor of number of citation is affected by h-index. To measure correlation of h-index toward number of citations this research using linear regression method because this method is very easy, quickly and can build with Excel, SPSS, Eviews, STATA, Minitab, MATLAB and manuals. and this research use Excel to build Multiple R to measure correlation level and R Square (R2) is coefficient determination to measure goodness of fit factors [17]. This research using confidence level 5%. The equation model of linear regression is following Equation 1.

$$y = a + by \qquad (1)$$

The values a and b can be calculated using Equation 2 and Equation 3.

$$a = \frac{(\Sigma y)\,(\Sigma x^2) - (\Sigma x)\,(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^{\,2}} \qquad (2)$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \tag{3}$$

The Equation 4 model to quantify correlation (r value or Multiple R) is

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{\{n\Sigma x^2 - (\Sigma x)^2\}\{n\Sigma y^2 - (\Sigma y)^2\}}} \tag{4}$$

The Equation 5 model to quantify r^2 or R Square is

$$r^2 = r \; x \; r \tag{5}$$

The Equation 6 model to quantify degree of freedom (df) is

$$df = n - 1 \tag{6}$$

The Equation 7 model to quantify Sum of Square (SS) is

$$SS = \Sigma(xi - \bar{x})^2 \tag{7}$$

The Equation 8 model to quantify Mean of Square (MS) is

$$MS = \frac{SS}{n - 1} \tag{8}$$

For another equation is according to the instructions on linear regression analysis [17].
Notes:
y = Response or Dependent Variable
x = Predictor or Independent Variable
a = Constanta
b = Regression Coefficient or scale of Response caused by predictor.
n = number of data observation

2.5 Output Data
This output data shown on Table 3 with regression statistics: Multiple R, R Square, Adjusted R Square, Standard Error and Observations.

Table 3. Regression Statistics

Regression Statistics	
Multiple R	0.78147853
R Square	0.610708693
Adjusted R Square	0.600464185
Standard Error	801.9448792
Observations	40

The test result shown that value of
1. Multiple R or r value= 0.78, is indicates the correlative is very closely,
2. R Square = 0.61, indicates coefficient determination the impact of h-index toward number of citation factor achieve 61% and the rest 39% affected by others factor.
3. Adjusted R Square = 0.60, indicates that the function of value is never decrease.
4. Standard Error 801.94, indicates that estimate dependent variable, value smaller standard error is compared to the standard deviation of the h-index, the more accurate the regression model in predicting the h-index.
5. Observation = 40 indicates that all data on S1 amount 40 journals.

Refer to Table 4, ANOVA (Analysis of Variance) is test the h-index acceptable from statistic perspective on analysis of the variant.

1. Degree of Freedom (df) from all data is n-1, where is "n" the is number of observations. The number of observation is 40, so number of total degrees of freedom is 39. Degree of freedom from this regression is 1, because only one degree of freedom (h-index). Degree of freedom for residual is the rest that total of degree of freedom reduce degree of freedom regression is 39-1=38.
2. Sum of Square (SS) obtained from addition square of dependent variable is subtracted with average value of h-index from real data, so form the manual we can find from original data. Then the result regression of calculation is 38338227.51. Residual SS column is obtained from square number from residual, then the result is 24438392.39. The result of total column of sum of square is 62776619.9.
3. Mean of Square (MS) is the result divided between the SS column with the df column. The regression MS number is 38338227.51 and residual MS is 643115.5893.
4. F is result from divided between MS regression with MS residual. This F value is known as F count in hypothesis testing compared to F table value. If F count > F table, it can be stated that h-index has significant effect to number of citation. In addition, we can also compare between the real levels with p-value (in Excel terms is Significance F). If the real level > p-value then the conclusion is the same as above. For example, we set a real level of 5%. Since p-value (Significance F) = 2.65, it can be concluded that h-index has a significant effect on the number of citations.

Table 4. Regression ANOVA

	df	SS	MS	F	Significance F
Regression	1	38338227.51	38338227.51	59.6132766	2.6596E-09
Residual	38	24438392.39	643115.5893		
Total	39	62776619.9			

Table 5. Regression Coefficients

	Coefficients	Standard Error	t Stat	P-value
Intercept	-486.2658519	227.3306579	-2.139024522	0.038921
H-index	140.7551548	18.23025788	7.72096345	2.66E-09

Refer to Table 5 *Regression Coefficient* is the impact of direction, the impact has two possibilities, that is positive and negative.
a = - 486.2658519 is value of Constanta.
b = 140.7551548 is regression coefficient that have negative direction, it means that if h-index is increase 140.7551548 than number of citation is decrease 140.7551548.
The Equation 9 of this regression is

$$y = a + bx \qquad (9)$$
$$y = -486.2658519 + 140.7551548\,x$$

3. Result and Discussion
Refer to interpretations of correlation coefficient, correlation and regression is very related closely, every regression must have correlation, but not every correlation has regression and this regression statistics are included regression with any correlation, so it can use to continue predict number of journal citation base on h-index at S1, see detail on Figure 4. The result takes from output data describe that the correlation between h-index and citation is included high correlation (Multiple R=0.78) and impact of h-index is to citation achieve 0.61 or 61 % while remaining 39% was influenced by other factors, and from equation of regression obtained that h-index is inversely proportional with citation, it means if h-index is increase 140 than citation is decrease 140 [20].

Refer to the output data and problem identification we can discuss that the journal citation is a key indicator of global acceptance, even on reputable indexing institution like Scopus and Web of Science, so recommended suggest to Kemenristekdikti that the accreditation system on Indonesian Journal (ARJUNA) add more value to the number of journal citations.
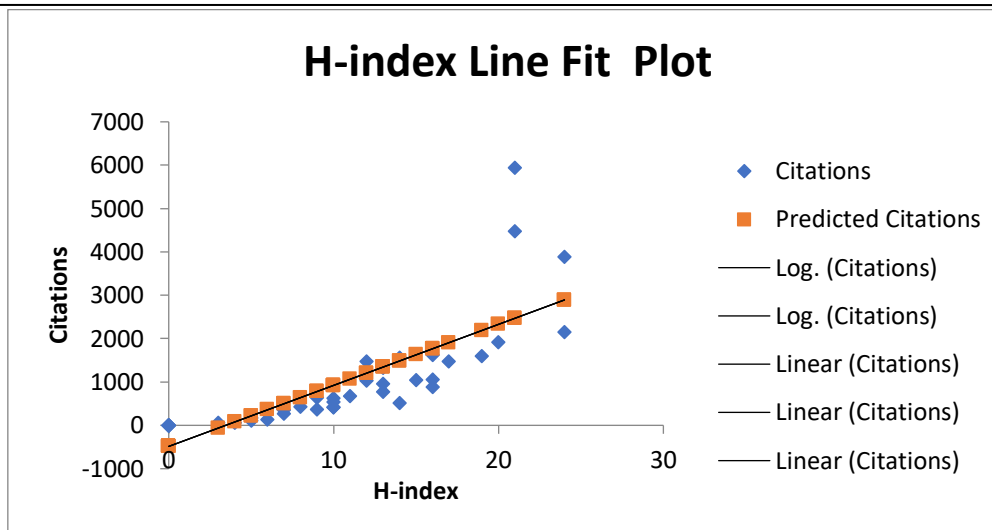
Figure 4. Predict of H-Index to Number of Journal Citation of 40 Journal on S1

4. Conclusion

This research focus to find how strong correlation or impact h-index toward citations using linear regression. The test result statistics on S1 shown that value of Multiple R = 0.78 is indicates the correlation between h-index and citation is included high correlation, value of R Square = 0.61 indicates the impact of h-index toward citations achieve 61% and the rest 39% affected by others factor.

References

[1] Presiden Republik Indonesia, *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi*. Jakarta, 2012, p. 97.

[2] Kemenristekdikti, *"Grafik Jumlah Dosen Aktif Pangkalan Data Pendidikan Tinggi,"* https://forlap.ristekdikti.go.id, 2018. .

[3] LIPI, *"Daftar ISSN yang telah diterbitkan,"* http://issn.lipi.go.id/, 2018. .

[4] Kemenristekdikti, *"SINTA - Science and Technology Index,"* http://sinta.ristekdikti.go.id, 2018. .

[5] C. Malesios, *"Measuring the Robustness of the Journal H-Index wth Respect to Publication and Citation Values: A Bayesian Sensitivity Analysis,"* Journal of Informetrics, vol. 10, No. 3, Pp. 719–731, 2016.

[6] L. Bertoli-Barsotti and T. Lando, *"On a formula for the h-index,"* Vol. 9, No. 4, Pp. 762–776, Elsevier Ltd, 2015.

[7] S. Asnafi, T. Gunderson, R. J. McDonald, and D. F. Kallmes, *"Association of h-index of Editorial Board Members and Impact Factor among Radiology Journals,"* Academic Radiology, Vol. 24, No. 2, Pp. 119–123, 2017.

[8] J. K. Vanclay, *"Ranking Forestry Journals Using the H-Index,"* Journal of Informetrics, Vol. 2, No. 4, Pp. 326–334, 2008.

[9] J. Mingers, F. Macri, and D. Petrovici, *"Using the H-Index to Measure the Quality of Journals in the Field of Business and Management,"* Information Processing and Management, Vol. 48, No. 2, Pp. 234–241, 2012.

[10] S. Chen, C. Arsenault, and V. Larivière, *"Are Top-Cited Papers More Interdisciplinary?,"* Journal of Informetrics, Vol. 9, No. 4, Pp. 1034–1046, 2015.

[11] V. Garousi and M. V Mäntylä, *"Citations, Research Topics and Active Countries in Software Engineering: A Bibliometrics Study,"* Computer Science Review, Vol. 19, Pp. 56–77, 2016.

[12] L. Waltman, *"A Review of the Literature on Citation Impact Indicators,"* Journal of Informetrics, Vol. 10, No. 2, Pp. 365–391, 2016.

[13] M. A. T. Lira, E. M. Da Silva, J. M. B. Alves, and G. V. O. Veras, *"Estimation of Wind Resources in the Coast of Ceará, Brazil, Using the Linear Regression Theory,"* Renewable and Sustainable Energy Reviews, Vol. 39, Pp. 509–529, 2014.

[14] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, *"Stock Market Predication Using a Linear

Regression," in 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Vol. 2, Pp. 510–513, 2017.

[15] Paikun, T. Kadri, and R. D. H. Sugara, *"Estimated Budget Construction Housing Using Linear Regression Model Easy and Fast Solutions Accurate,"* in 2017 International Conference on Computing, Engineering, and Design (ICCED), Pp. 1–6, 2017.

[16] S. P. Menon, R. Bharadwaj, P. Shetty, P. Sanu, and S. Nagendra, *"Prediction of Temperature Using Linear Regression,"* in 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Pp. 1–6, 2017.

[17] G. G. V. Douglas C. Montgomery, Elizabeth A. Peck, *"Introduction to Linear Regression Analysis, 5th Edition,"* Fifth. Canada: WILEY, 2012.

[18] D. Muriyatmoko, *"Editorial Team,"* https://ejournal.unida.gontor.ac.id/index.php/FIJ, 2018. .

[19] Scopus, *"SJR : Scientific Journal Rankings,"* Scimago Journal & Country Rank, 2018. .

[20] J. P. Guilford, *"Fundamental Statistics in Psychology and Education,"* 5th Edition. New York [etc.]: McGraw-Hill Book Company, 1973.