



The Research Trend of Statistical Significance Test: Bibliometric Analysis

Dwi Fitria Al Husaeni ¹, Dwi Novia Al Husaeni ¹, Meli Fiandini ², Asep Bayu Dani Nandiyanto ^{3,*}

¹ Departemen Pendidikan Ilmu Komputer, Universitas Pendidikan Indonesia, Indonesia

² Departemen Pendidikan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia, Indonesia

³ Departemen Kimia, Universitas Pendidikan Indoensia, Indonesia

*Correspondence: E-mail: nandiyanto@upi.edu

ABSTRACT

A significance test is a test that can determine whether a research hypothesis is accepted or not. Therefore, significance tests need to be carried out, especially for researchers who will conduct research related to hypothesis testing. This research discusses publication trends regarding significance tests. This research aims to analyze research trends regarding statistical significance tests using bibliometric analysis methods assisted by mapping analysis. Bibliometric analysis methods and mapping analysis with VOSviewer were used in this research. The data source was taken from the Google Scholar database in the period 2013 - 2023. The total number of documents found during the last 10 years was 1000 documents. Based on the research results, it was found that the number of publications regarding statistical significance tests has decreased every year. From a total of 281 documents in 2013 to 7 documents in 2023. The results of article data mapping produce three forms of visualization, namely network, overlay, and density visualization. The mapping results show that research on statistics still needs to be carried out, especially regarding significance tests. This research concluded that the trend of research discussing statistical significance tests is decreasing, although in terms of its use, it is still very important, especially in processing quantitative research data.

ARTICLE INFO

Article History:

Submitted/Received 07 Jun 2023

First Revised 25 Jul 2023

Accepted 21 Sep 2023

First Available online 22 Sep 2023

Publication Date 03 Mar 2024

Keyword:

Bibliometric analysis,
Significance test,
Statistics.

1. INTRODUCTION

Significance testing is one of the most important stages in research, especially in research that uses a quantitative data processing approach. The significance test determines whether a research hypothesis is accepted or not. Significance tests help researchers determine conclusions from research results (Onwuegbuzie & Leech, 2004). Statistical significance is a sufficient level of confidence to be able to accept a hypothesis that has been determined in a study. In research, Cox (1982) said that the first element in a significance test is the formulation of the hypothesis to be tested, often called the null hypothesis. Significance tests relate to several aspects of assessing the information provided by an investigation (Cox, 1982).

Many previous studies used significance tests, including the redefinition of statistical significance tests (Benjamin et al., 2018), discussion of scientists' opposition to statistical significance (Amrhein et al., 2019), discussion of theoretical analysis of P values and statistical significance (Andrade, 2019), research looking at current trends in the use of statistical tests to compare cluster and evolutionary computing algorithms with a critical review (Carrasco et al., 2020), and research on p-value based statistical significance tests (Zhang, 2022).

Based on several previous studies regarding significance test statistics, there has been no research that analyzes research trends regarding significance test statistics using a bibliometric analysis approach assisted by mapping visualization. Recently, the use of bibliometric analysis has often been used by many previous studies in determining research trends (Nandiyanto et al., 2023; Ragadhita & Nandiyanto, 2022; Kim et al., 2021; Suprpto et al., 2021; Niknejad et al., 2021). **Table 1** shows several previous studies that used bibliometric analysis to determine trends or developments in a research theme in more detail. Therefore, this research aims to analyze research trends regarding statistical significance tests using bibliometric analysis methods assisted by mapping analysis. This research also analyzes the development of research and the development of the number of citations in research regarding statistical significance tests. We use visualization mapping to determine discussion keywords in research regarding statistical significance tests.

Table 1. Previous research discussing research trends using bibliometric analysis.

No	Research Subject	Data source	Year Range	Application	Ref
1	Salicylic Acid	Scopus	2000-2021	VOSviewer	Ruzmetov and Ibragimov (2023)
2	Flooding Stress in Plant Science and Agriculture Subject Areas	PubMed, Google Scholar, Semantic Scholar, and Crossref	1962-2021	VOSviewer	Nurrahma et al. (2023)
3	Water Hyacinth and Education	Scopus	2010-2023	Microsoft Excel, VOSviewer	Hofifah and Nandiyanto (2024)
4	Water Hyacinth and Ecosystem	Scopus	2000-2023	VOSviewer	Nandiyanto et al. (2024)
5	Briquette	Google Scholar	2017-2022	VOSviewer, Microsoft Excel	Al Husaeni (2022)
6	Management Bioenergy	Google Scholar	2017-2021	VOSviewer	Soegoto et al. (2022)
7	Philosophy of science and technology education	Scopus	2018-2022	VOSviewer	Al Husaeni & Munir (2023)

2. METHODS

Bibliometric analysis was carried out as a research method. There are several stages of bibliometric analysis, including: Collecting article data is the first step in conducting literature research using bibliometric analysis. At this stage, published research documents related to the topic "Statistical significance" are collected. The article data used is article data indexed by Google Scholar from 2013 - 2023. Article data is collected via the Publish or Perish application. The results of collecting article data using Publish or Perish resulted in 1000 articles for analysis. The collected research article data is saved in (*.csv) format so that it can be analyzed using Microsoft Excel software, and (*.ris) format so that it can be analyzed and visualized using the VOSviewer application. After data collection, article data was filtered to see the completeness of components (such as year). Next, the article data was analyzed using Ms. Excel and visualized using VOSviewer. More detailed stages regarding the analysis steps are explained in our previous research ([Al Husaeni & Nandiyanto, 2022](#)).

3. RESULTS AND DISCUSSION

3.1. Development of statistical significance test publications 2013-2023

Table 2 shows the annual report on research on "Statistical significance" which has been published in national and international journals. Based on the data, it is known that the total number of documents found over the last 10 years is 1,000 documents. Details of the number of research documents regarding "Statistical significance" namely 2013 as many as 281 documents, 2014 as many as 203 documents, 2015 as many as 176 documents, 2016 as many as 107 documents, 2017 as many as 72 documents, 2018 as many as 55 documents, 2019 as many as 26 documents, 2020 as many as 49 documents, 2021 there will be 16 documents, and 2023 there will be 7 documents.

Table 2. Annual Report Research on "Statistical significance".

Year	Documents	Percentages (%)
2013	281	28.10
2014	203	20.30
2015	176	17.60
2016	107	10.70
2017	72	7.20
2018	55	5.50
2019	26	2.60
2020	49	4.90
2021	16	1.60
2022	8	0.80
2023	7	0.70
Total	1000	100

Based on the number of research documents each year, it is known that research publications regarding "Statistical significance" have decreased from 2013 to 2023. **Figure 1** shows a graph of the decline in the number of publications regarding "Statistical significance" more clearly. Over the last 10 years, the highest number of studies on this topic was in 2013 (281 documents) and the lowest number was in 2023 (7 documents). The decrease in the number of documents occurs consistently every year, but there was an increase of around 23 documents from 2019 to 2020. However, in 2021 the number of published documents on this topic experienced a significant decrease of around 33 documents.

Statistics are very important for everyone to know. Statistics can help researchers to conclude that the differences obtained are truly significantly different. Apart from that, the use of statistics is very broad and covers almost all areas of human life, from development, economics, and health, to analysis of human behavior. In the field of economics, statistics are very important in explaining economic growth, inflation, and business development. Therefore, this research was conducted to describe the literature review process regarding statistical calculations for various purposes.

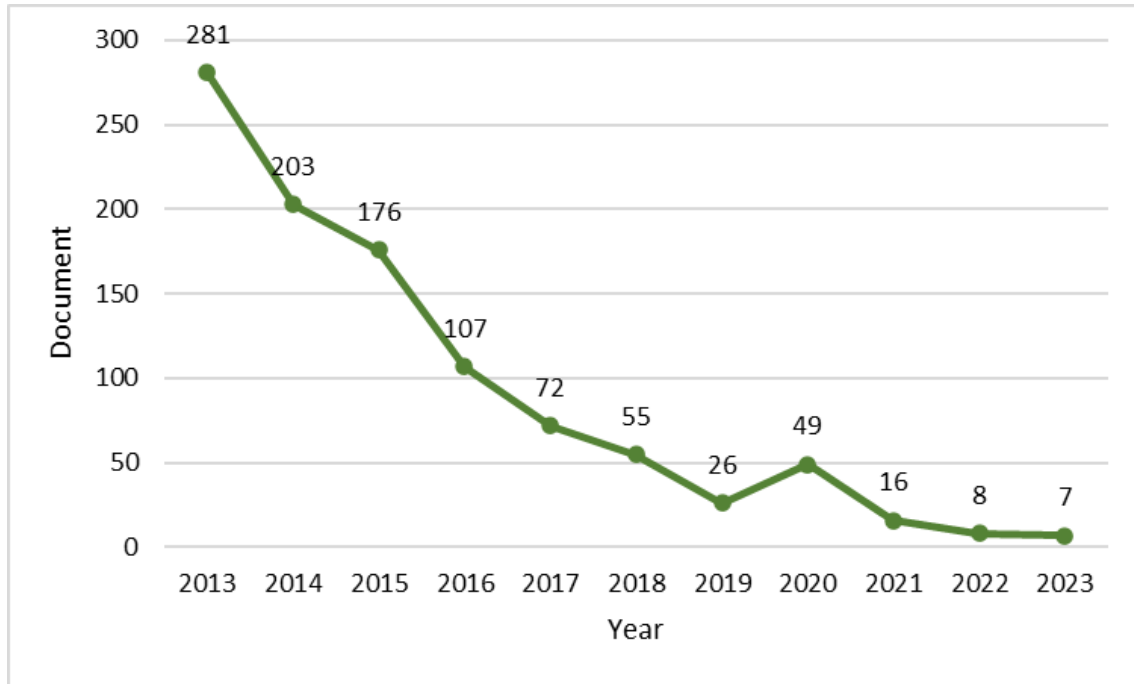


Figure 1. Research Developments regarding "Statistical significance"

3.2. Trend of Statistical Significance Test Research Citations 2013-2023

In this research, we present 10 articles regarding different test statistics that have the highest number of citations. **Table 3** presents some metadata from articles with the highest number of citations. Based on **Table 3**, it is known that many articles with the title "lmerTest package: tests in linear mixed effects models" written by Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. in 2017 are articles about difference test statistics that are most frequently cited. total 17222 cited. [Kuznetsova et al. \(2017\)](#) discussed a literature review regarding the lmerTest package. In addition, [Kuznetsova et al. \(2017\)](#) applied Satterthwaite's method to estimate degrees of freedom for t and F tests and construction of Type I–III ANOVA tables. The other article with the largest number is research conducted by [Zinman et al. \(2015\)](#). [Zinman et al. \(2015\)](#) discussed statistical calculations regarding Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. The article written by [Zinman et al. \(2015\)](#) has been cited 11461 times since 2015, with the average number of citations per year being 1432.63 times. Based on the data in **Table 3**, it is known that published articles regarding different test statistics with the highest number of citations came from 7 different countries, namely Denmark (1 document), New England (2 documents), USA (3 documents), United Kingdom (1 document), China (1 document), Turkey (1 document), and Spain (1 document).

Table 3. Difference Test Statistics articles with the most citations.

No	Cites	Authors	Title	Year	Cites PerYear	Country	Ref
1	17222	Kuznetsova <i>et al.</i>	lmerTest package: tests in linear mixed effects models	2017	2870.33	Denmark	Kuznetsova <i>et al.</i> (2017)
2	11461	Zinman <i>et al.</i>	Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes	2015	1432.63	New England	Zinman <i>et al.</i> (2015)
3	6984	Neal <i>et al.</i>	Canagliflozin and cardiovascular and renal events in type 2 diabetes	2017	1164.00	New England	Neal <i>et al.</i> (2017)
4	6239	Hinshaw <i>et al.</i>	Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results	2013	623.90	USA	Hinshaw <i>et al.</i> (2013)
5	6069	Everingham <i>et al.</i>	The Pascal Visual Object Classes Challenge: A Retrospective	2015	758.63	United Kingdom	Everingham <i>et al.</i> (2015)
6	5749	RV Lenth	Least-squares means: the R package lsmeans	2016	821.29	USA	Lenth (2016)
7	4912	Chen <i>et al.</i>	Clinical and immunological features of severe and moderate coronavirus disease 2019	2020	1637.33	China	Chen <i>et al.</i> (2020)
8	4006	Blevins <i>et al.</i>	The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation	2015	500.75	USA	Blevins <i>et al.</i> (2015)
9	3915	H Akoglu	User's guide to correlation coefficients	2018	783.00	Turkey	Akoglu (2018)
10	3751	Fernández-Delgado <i>et al.</i>	Do we need hundreds of classifiers to solve real world classification problems?	2014	416.78	Spain	Fernández-Delgado (2014)

3.3. Visualization of Research Data Mapping of Statistical Significance Test Research

Data mapped using VOSviewer produces 3 forms of visualization, namely network visualization (**Figure 2**), overlay visualization (**Figure 3**), and density visualization (**Figure 4**). Network visualization shows that the terms generated from the abstract and keywords that are considered to correspond to the keywords used when collecting data are divided into 6 clusters with a total of 96 items. Each item has a different link, total link strength, and occurrences. Overall, based on network visualization, the total link strength is 12133 while the total number of links is 2625. The following is a more detailed explanation of each cluster:

(i) Cluster 1 marked in red consists of 26 items, namely average, case, coefficient, control group, correlation, data, difference, evidence, experiment, firm, impact, mean, model, normal distribution, performance, relationship, sample, significant difference, statistics, statistical significance, t test, table, test, value, and variable.

- (ii) Cluster 2 marked in green consists of 26 items, namely ages, association, change, comparison, covid, day, disease, group difference, mean age, mean change, month, normal range, outcome, p value, patient, placebo, pregnancy, risk, score, study, testing, time, treatment, trial, week, and year.
- (iii) Cluster 3 marked in blue consists of 21 items, namely diabetes, effect, effect size, efficacy, estimate. Heterogeneity, intervention, meta-analysis, metaanalysis, null hypothesis, number, randomized controlled trial, significant effect, SMD, standard deviation, standardized mean difference, standardized mean difference, systematic review, type, weighted mean difference, and woman.
- (iv) Cluster 4 marked in yellow consists of 15 items, namely analysis, average difference, gender, hypothesis, influence, level, mean difference, measure, person, power, reliability, significance, significant result, and statistical test.
- (v) Cluster 5 marked in purple consists of 6 items, namely depression, group, mean score, significance level, student, and variance.
- (vi) Cluster 6 marked in light blue consists of 2 items, namely evaluation and statistical analysis.

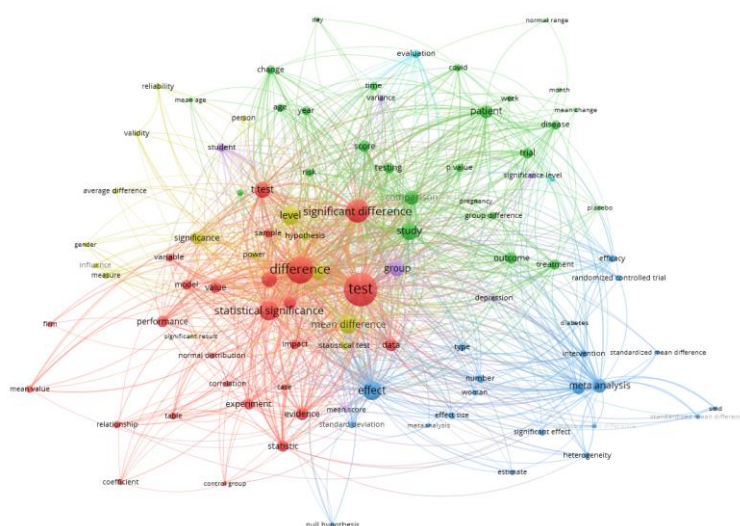


Figure 2. Network visualization based on co-occurrence of terms.

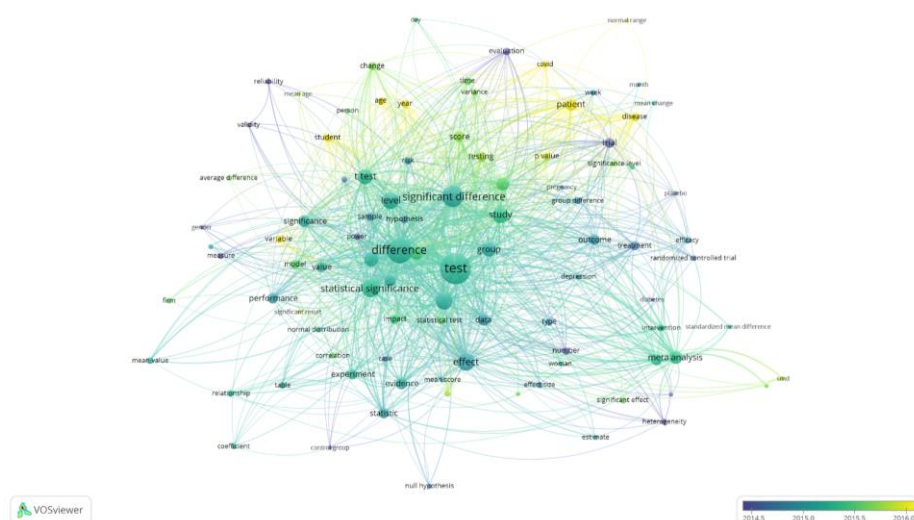


Figure 3. Overlay visualization based on co-occurrence of terms.

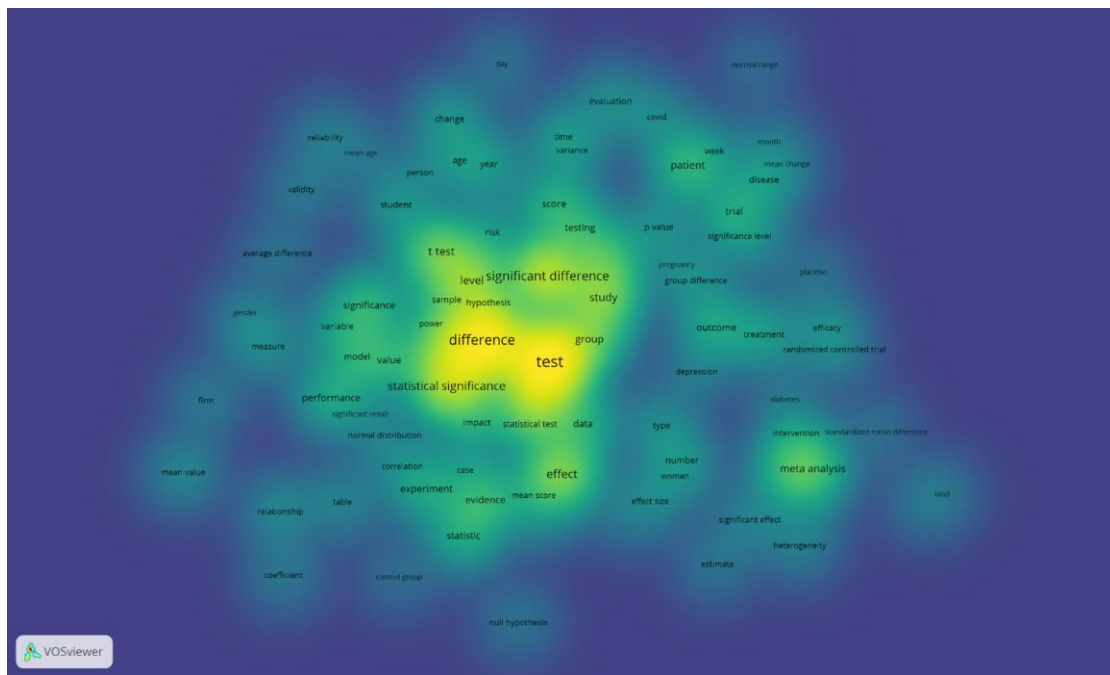


Figure 4. Density visualization based on co-occurrence of terms.

Apart from that, based on network visualization (**Figure 2**), the terms used as keywords are in cluster 1, such as test which has 520 occurrences with a total link strength of 2230, and different which has 361 occurrences with a total link strength of 1062, and statistical significance which has occurrences were 185 with a total link strength of 859 (see **Figure 5**).

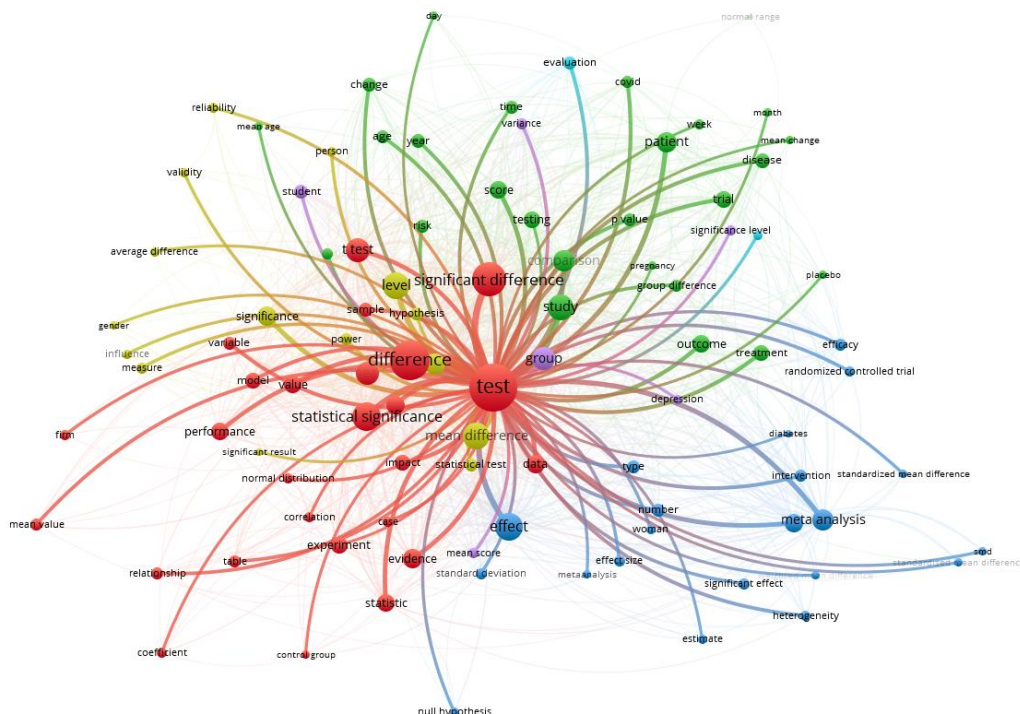


Figure 5. Network visualization of Cluster 1.

If in the network visualization, we can see the division of clusters, the connection between items and items that are appropriate and considered relevant to the keywords used to search

for articles, then in the overlay visualization (**Figure 3**), we can see the time span (years) for each term studied. For example, the term test, if you look at the overlay visualization, the term test was widely used in 2015 (marked in green), while the term control group was widely used in 2014 (marked in purple), and the term student was widely used in 2016 (marked in yellow). In contrast to network and overlay visualization, density visualization (**Figure 4**) shows the most frequently used terms using color, meaning that the brighter the color of a term, the more often that term is used.

4. CONCLUSION

The number of publications regarding "Statistical significance" during the last 10 years (2013-2023) was 1000 documents. The development of research regarding "Statistical significance" has decreased. 2013 was the year with the highest number of publications, namely 281 articles. This number continued to decrease until 2019 with the respective numbers being 203 (2014), 176 (2015), 107 (2016), 72 (2017), 55 (2018) and 26 (2019). In 2020, the number of publications increased to 49 documents. This increase did not last long because, in the following year, it decreased again to 16 documents (2021), 8 documents (2022), and 7 documents (2023). Apart from that, the terms used as keywords when used in the mapping stage using the VOSviewer application produce 96 items which are divided into 6 clusters marked with different colors. Each cluster has a different number of items and each item has a different number of occurrences, links, and total link strength (although it does not rule out the possibility that there are the same number of items). Terms that are often used both as keywords and as abstracts in research articles related to keywords such as test, difference, and statistical significance (widely used in 2015).

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

7. REFERENCES

- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3), 91-93.
- Al Husaeni, D. F., and Nandiyanto, A. B. D. (2022). Bibliometric using Vosviewer with Publish or Perish (using google scholar data): From step-by-step processing for users to the practical examples in the analysis of digital learning articles in pre and post Covid-19 pandemic. *ASEAN Journal of Science and Engineering*, 2(1), 19-46.
- Al Husaeni, D.F., and Munir, M. (2023). Literature review and bibliometric mapping analysis: Philosophy of science and technology education. *Indonesian Journal of Multidiciplinary Research*, 3(2), 219-234
- Al Husaeni, D.N. (2022). Bibliometric analysis of briquette research trends during the covid-19 pandemic. *ASEAN Journal for Science and Engineering in Materials*, 1(2), 99-106.
- Amrhein, V., Greenland, S., and McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305-307.
- Andrade, C. (2019). The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian Journal of Psychological Medicine*, 41(3), 210-215.

- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., and Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10.
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., and Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, 28(6), 489-498.
- Carrasco, J., García, S., Rueda, M. M., Das, S., and Herrera, F. (2020). Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review. *Swarm and Evolutionary Computation*, 54, 100665.
- Chen, G., Wu, D. I., Guo, W., Cao, Y., Huang, D., Wang, H., Wang, t., Zhang, X., Chen, H., Yu, h., Zhang, X., Zhang, M., Wu, S., Song, J., Chen, T., Meifang, H., Li, S., Luo, X., Zhao, J., and Ning, Q. (2020). Clinical and immunological features of severe and moderate coronavirus disease 2019. *The Journal of Clinical Investigation*, 130(5), 2620-2629.
- Cox, D. R. (1982). Statistical significance tests. *British Journal of Clinical Pharmacology*, 14(3), 325-331.
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111, 98-136.
- Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. *The Journal of Machine Learning Research*, 15(1), 3133-3181.
- Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C., Dunkley, J., Nolte, m. R., Halpern, M., Hill, R. S., and Wright, E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2), 19.
- Hofifah, S. N., and Nandiyanto, A. B. D. (2024). Water hyacinth and education research trends from the scopus database: A bibliometric literature review. *ASEAN Journal of Science and Engineering Education*, 4(2), 121-132.
- Kim, B. J., Jeong, S., and Chung, J. B. (2021). Research trends in vulnerability studies from 2000 to 2019: Findings from a bibliometric analysis. *International Journal of Disaster Risk Reduction*, 56, 102141.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1-26.
- Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69, 1-33.
- Nandiyanto, A. B. D., Fiandini, M., and Al Husaeni, D. N. (2024). Research trends from the scopus database using keyword water hyacinth and ecosystem: A bibliometric literature review. *ASEAN Journal of Science and Engineering*, 4(1), 33-48.

- Nandiyanto, A. B. D., Ragadhita, R., Al Husaeni, D. N., and Nugraha, W. C. (2023). Research trend on the use of mercury in gold mining: Literature review and bibliometric analysis. *Moroccan Journal of Chemistry*, 11(1), 11-1.
- Neal, B., Perkovic, V., Mahaffey, K. W., De Zeeuw, D., Fulcher, G., Erondy, N., Shaw, W., Law, G., Desai, M., and Matthews, D. R. (2017). Canagliflozin and cardiovascular and renal events in type 2 diabetes. *New England Journal of Medicine*, 377(7), 644-657.
- Niknejad, N., Ismail, W., Bahari, M., Hendradi, R., and Salleh, A. Z. (2021). Mapping the research trends on blockchain technology in food and agriculture industry: A bibliometric analysis. *Environmental Technology and Innovation*, 21, 101272.
- Nurrahma, A. H. I., Putri, H. H., and Syahadat, R. M. (2023). Scientific research trends of flooding stress in plant science and agriculture subject areas (1962-2021). *ASEAN Journal of Science and Engineering*, 3(2), 163-178.
- Onwuegbuzie, A. J., and Leech, N. L. (2004). Enhancing the interpretation of "significant" findings: The role of mixed methods research. *The Qualitative Report*, 9(4), 770-792.
- Ragadhita, R., and Nandiyanto, A. B. D. (2022). Computational bibliometric analysis on publication of techno-economic education. *Indonesian Journal of Multidisciplinary Research*, 2(1), 213-222.
- Ruzmetov, A., and Ibragimov, A. (2023). Past, current and future trends of salicylic acid and its derivatives: A bibliometric review of papers from the Scopus database published from 2000 to 2021. *ASEAN Journal for Science and Engineering in Materials*, 2(1), 53-68.
- Soegoto, H., Soegoto, E.S., Luckyardi, S., and Rafdhi, A.A. (2022). A bibliometric analysis of management bioenergy research using vosviewer application. *Indonesian Journal of Science and Technology*, 7(1), 89-104.
- Suprpto, N., Sukarmin, S., Puspitawati, R. P., Erman, E., Savitri, D., Ku, C. H., and Mubarak, H. (2021). Research Trend on TPACK through Bibliometric Analysis (2015-2019). *International Journal of Evaluation and Research in Education*, 10(4), 1375-1385.
- Zhang, W. (2022). p-value based statistical significance tests: Concepts, misuses, critiques, solutions and beyond. *Computational Ecology and Software*, 12(3), 80-122.
- Zinman, B., Wanner, C., Lachin, J. M., Fitchett, D., Bluhmki, E., Hantel, S., Mattheus, M., Devins, T., Johansen, O. E./ Woerle, H. J., broedl, U. C., and Inzucchi, S. E. (2015). Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *New England Journal of Medicine*, 373(22), 2117-2128.