Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 5, October 2025, pp. 4138~4149
ISSN: 2302-9285, DOI: 10.11591/eei.v14i5.9506

New approach to measuring researcher expertise using cosine similarity algorithm and association rules

Ali Firdaus[1,2], Deris Stiawan[2], Samsuryadi[2], Rahmat Budiarto[3]
[1]Department of Computer Engineering, Faculty of Computer Engineering, Politeknik Negeri Sriwijaya, Palembang, Indonesia
[2]Department of Computer Science, Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia
[3]Department of Computer Engineering and Science, College of Computer Science and IT, Al-Bahah University, Al Aqiq, Saudi Arabia

Article Info

Article history:

Received Nov 3, 2024
Revised Aug 15, 2025
Accepted Sep 1, 2025

Keywords:

Collaboration
Experts
Measurement
Quality
Ranking
Scientific

ABSTRACT

This study proposes a new method to assess researcher expertise using publication data. The quality of research publications is an important indicator in the ranking of universities that are undergoing diversification. Research publications have become an important indicator in the university ranking system and have a major impact on the reputation of universities as a lens for the study of expertise and prestige for human resources. Expertise is often difficult to verify objectively, as a result, many people claim to be experts or are considered experts without evidence and correct data. To ensure the expertise of researchers, it must be proven with valid data support through measurable and presentable expertise parameters. The model built uses the cosine similarity and association rule approaches. The publication variables attached to the researcher are formulated in the collaboration of the algorithm to assess the level of researcher expertise. Validation of important points of publications as parameters for measuring expertise has been identified as the main factor contributing to the measurement of researcher expertise and its impact on university reputation. The model built successfully validated researcher expertise up to 72% which is relevant to its support for university rankings up to 75%.

Corresponding Author:

Deris Stiawan
Department of Computer Science, Faculty of Computer Science, Sriwijaya University
Big Hill Area, Srijaya Road, Palembang, Indonesia
Email: deris@unsri.ac.id

1. INTRODUCTION

The ranking system in higher education is adopted as an instrument that provides information and evaluation guidelines at the global level [1], [2]. It was initiated by the Institute of Higher Education of Shanghai Jiao Tong University in 2003 [3]. Times Higher Education World University Ranking allocates 30% of the total ranking value for the number of citations [4], [5]. So publications and citations are important for universities to maintain their position in the competition [6].

The success of an organization is built on expertise as a combination of knowledge and years of experience [7]. Knowledge is a fundamental and important component for the sustainability of higher education that is consistent in creating new knowledge, disseminating it to many organizations and appreciating it is a success [8], [9] those who ignore this role will not be appreciated [10]. The publication trail can be recognized using measurable publication variable parameters [11], [12]. The quality of research publications determines the difference in the reputation of a university as the foundation of modern science [13], by producing knowledge, publications, citations in highly indexed journals [14].

There are four important aspects, namely; facts, developments, content, and research recognition [15], [16]. Science defines, describes, and explains phenomena, research efforts, and results [17]. The development of skills and knowledge is needed in an ever-changing and highly competitive environment [18], [19]. Information technology architecture supports the utilization and development of knowledge for higher education innovation as a strategic basis for maintaining competitive advantage [20], [21].

Expert knowledge is represented in research for investigation, revision, comparison, and discovery of the latest facts by proposing new theories and laws or for applying the knowledge [22], [23] which is integrated as an important component in solving complex problems [24].

Expertise is the fuel and weapon in success with scientific support to help, facilitate, and encourage performance in higher education in preparing data, analysis, formulating problems, providing intellectual weight to the issues and problems being discussed [25]. Involving experts ensures the quality of decisions taken [26]. Expertise and knowledge are intangible power motors that cannot be measured with money [26]-[28].

Expertise is currently not well measured, many people are considered or claim to be experts only assessed from the field of work that they have been working on for a long time, or because they often appear in the media as certain speakers without looking at their knowledge objectively. On the other hand, Wilkins *et al*. [29] explained that many experts are negligent in maintaining an adequate self-description of their ever-changing specialist expertise. This study contributes a new approach to the method of measuring researcher expertise using the cosine similarity and association rule algorithms with variable originality. This study is a development of previous research contributions that only focused on measuring researcher expertise [22].

This study combines the cosine similarity algorithm and association ruler in calculating expertise scores, by extending the analysis of publication contributions to university rankings and competitiveness. Section 2 of this study discusses the current theory of expertise assessment. Section 3 explains the work plan through research questions. Section 4 discusses the results achieved. Finally, section 5 writes the results achieved as a basis for future work.

2. RELATED WORK

The studies in [30], [31] discuss expert search that ignores contextual knowledge and only considers the work environment and peers or their work ability. Hofmann *et al.* [32] presents contextual expert search, which explores the data of work ability and context, and the problem of data mining, in searching; i) expertise ranking records or ii) expert ranking data tables according to query rules.

He *et al*. [33] investigate how to combine and assess the suitability of diverse knowledge to improve the expert search process. A discriminative probabilistic model proposed by Ferguson *et al*. [34] has characterized the latent topics and graph levels for those who match the profiles. Sharma *et al*. [35] proposed a combinational reference personalization on structured semantic information available in the research community that leverages encyclopedic knowledge sources and large news article datasets. A friendship strength recommendation system on topics or interests that requires users to analyze data on Twitter and transform features, items, user words into the same vector space to measure how much users like an item based on the highlighted item interest and polarity aspects has been proposed [36]-[38]. Association rules have access to generate relationships between large sets of data through navigation paths. The large number of data sets causes redundancy so that recommendations are inaccurate [39], [40]. In measuring expertise, several approaches have been taken to analyze the relationship and strength of variables as parameters. Table 1 explains several research sources for expert search with several algorithms.

Finding methods for measuring and providing information on expertise needs to be explored and developed through indicators that are acceptable and represent expertise in various fields of research ethics and integrity to support the comprehensive search and measurement of researcher expertise and assess its impact on university performance and ranking.

Social network content has been used as a model to measure user similarity in expert search by exploiting their social relationship graph in sharing common interests such as friendship, sharing, liking, following, and commenting on social media. On the other hand, to combine multiple expertise estimators and information extracted from users in expert search derived from textual similarity between documents and queries is done by exploiting the rank aggregation method as revealed in previous research [41]-[44]. In another part, the cognitive model also becomes an authentic solution in certain scientific disciplines empirically in the study of skilled practitioners ('experts') [45].

From the observation of previous research reviews, it is necessary to work hard to model correctly the users in general to find and measure expertise because there are still weaknesses in previous approaches such as systems that are difficult to represent and compare users on social media who do not like to follow or be followed, due to the variety of expertise and systems that are able to follow and are difficult to adapt to different cases. There is a high possibility of failure for independent and closed users. So, there must be a new approach in the field of researcher expertise with more measurable methods and variables.

Table 1. Relevant previous research

No	Authors	Parameter	Method
1	Mizzaro *et al*. [41]	Computes similarity based on semantic relation network between words occurring in the same tweet and related topics. Users are represented by posts of words.	Using a special network to determine multiple user profiles to compare by cosine similarity.
2	Pavan *et al*. [42]	Presenting a first attempt to create an expert search system to support users (researchers, students, and writers) in finding experts.	Expert search is based on community knowledge graph for semantic enrichment. Dynamic graph is built during user set analysis and allows to determine user similarity by comparing terms and entities.
3	Noureddine *et al*. [43]	A principled approach to combining multiple expertise estimators, derived from textual content, from graph-structured citation patterns for expert communities, and from profile information about experts.	We exploit unsupervised ranking aggregation methods to combine multiple skill estimators and information extracted from user profiles. We specifically experiment with two approaches, namely CombSUM, CombMNZ.
4	Birtolo *et al*. [44]	Exploiting ontology and structured and unstructured data from multiple web sources with different content. Correlation-based profiling, by exploiting heterogeneous web sources.	Correlation-based approach to researcher profiling: CARP. The aim is to address data quality issues and provide comprehensive and validated information about researchers and experts in the computer science domain, bypassing the matching procedure.
5	Price *et al*. [45]	Assess data sets and evaluate the level of data sparsit	Proposing a model-based collaborative filtering algorithm.
6	Barrantes and Ortiz [46]	Identifying survey papers related to the theme of text mining and semantics, and research, identified in digital libraries.	Latent semantic indexing (LSI) is a method that can be used for data dimension reduction and is known as latent semantic analysis.
7	Maharana *et al*. [47]	Identifying the top 23 journal publications in Medical Informatics based on the Institute for Scientific Information's (ISI) web of knowledge Journal Citation Report (JCR).	Researchers took article summaries from selected samples from the PubMed website. The documents were then transformed and arranged into a form of that was similar to that of a computer process.
8	Amatulli *et al*. [48]	Identify positive, negative and neutral opinions from text. sources of opinion personal blogs and online review sites.	Pre-processing, data is cleaned with parts of speech (PoS) and threshold-based data partitioning (TDP) algorithm. Merging clusters into one cluster.
9	Braun *et al*. [49]	Explore and develop acceptable indicators that represent expertise in various areas of research ethics and integrity.	The qualitative study was complemented by a quantitative survey among a broader group of practitioners in research ethics and integrity and participatory research with a series of consensus conferences involving assumed users of the expert.
10	Wu *et al*. [50]	Analyzing publication preferences and calculating the similarity of research preferences over a 5-year period, as measured by the relative percentage of publications in different subject areas. This analysis selects the twenty academic institutions that ranked with the highest scores in the latest edition of THE World University Rankings 2020, and applies bibliometric indicators from Elsevier's SciVal [50].	Compare and extract publication data from Scopus, apply bibliometric indicators, cosine similarity statistical techniques and agglomerative hierarchical clustering analysis to examine, compare research preference affinities. Cluster analysis through VOSviewer to classify total scientific production in health, physical, life, and social sciences.
11	Wei *et al*. [51]	Extensive on problem solving and expertise, teaching and assessing advanced problem solving skills in post-secondary students applicable across the disciplines of science, engineering, and medicine [51].	Based on a cognitive model of the problem-solving process that is empirically grounded in studies of skilled practitioners ('experts') solving authentic problems in their disciplines.

Research and publication expertise has a strong influence on university rankings and achievements. Universities have extensive dedication and recommendations in research and are widely recognized for their research expertise and establishment. The classification of Higher Education Institutions in developed countries has shown a significant influence on research and publications on the image and reputation of universities in modern world thought and has a great influence on society and culture throughout the world [45]. Some research institutions have proven to be the strongest universities and have produced many international discoveries. Some are even so specialized that they are used as references as strong scientific foundations such as the University of Oxford in medical studies [46].

2.1. Research gap

The main research gap from previous research is that no measurement of researcher expertise is found through publication variables as a measurable parameter that can be scientifically proven. There are five gaps in previous research, as depicted in Table 2.

Table 2. The research gap

No	Research gap	Differences and similarities in research
1.	Evidence	In the previous research above, it only tried to find experts through their relationships with social networks, profiles, knowledge, positions, affiliations, publication rankings but did not measure how competent a person's expertise is, and there are still variables that are difficult to measure such as "Calculating user similarity based on the semantic relationship network between words that occur in the same tweet and related topics, users are represented by frequently posted words, and other words from the text enrichment procedure, with the tweets they make, and the topics that users are interested in [52]. Domain-dependent independent knowledge, extracted from user-published content and representing the user's level of domain knowledge for a given query, and derived from network structure and social information to determine the user's level of expertise [53].
2.	Knowledge	The knowledge gap in this study with previous studies is; existing knowledge for theory and literature from the research domain related to the case is different and does not focus on measuring researcher expertise, the research results are different from what is expected.
3.	Methodological	The research methods used in previous studies have not been measured well because they still involve the influence of social media users to obtain input in the measurement, so a new approach is needed.
4.	Theoretical	This theoretical gap is related to the theory used in previous research. Previous research has not explained the theory of measuring a person's expertise, some theoretical models used to explain there are still theoretical conflicts so that there needs to be renewal.
5.	Population	Population gaps and lack of focus on researchers' publication results.

3. METHOD

The flow of the research plan using a renewable approach in measuring expertise with algorithmic collaboration can be seen in Figure 1. In the measurement as presented in Figure 1 it explains that in measuring expertise the initial step that must be taken is to assess the similarity of the publication title to the data dictionary in a particular field of science. Then predict the percentage of similarity, if it meets the criteria, it will be processed and collaborated with the association rulers measure to predict the value of the researcher's expertise. Furthermore, the steps for measuring the researcher's expertise start from answering the research question, then determining the measurement method to get results as in the Figure 1.

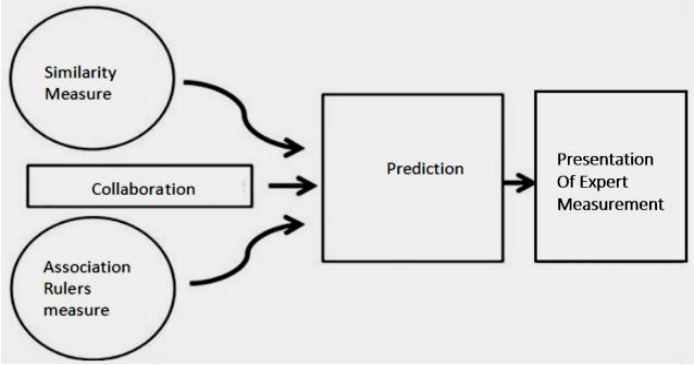

Figure 1. Research design with algorithmic collaboration

Figure 2 explains the steps of this research, starting with answering the research questions to achieve the research objectives. Next, analyze the variables to be processed and determine the data processing method, to group research in the same direction of focus and display the percentage of expertise. Here, we apply a new approach rule for the extraction of similarity measure features. This methodology is illustrated in Figure 3. The computational framework is as follows; i) initial data (data standard words), ii) feature determination (searching for publication data similarities), iii) association rules on feature determination to get weights for the most genes, and iv) pruning and grouping association rules based on manipulation of similarity measures, and displaying expertise values.

The word phrase extractor identifies word phrases in the text. The word phrase indexer is responsible for indexing word phrases found in the document. The normalized term frequency score (tf score) is used to select prominent word phrases. The association rule miner extracts association rules between word phrases found by the previous component.

The research data were taken randomly from the SINTA domain ("https://sinta.kemdikbud.go.id") which provides access to citations and expertise to measure researcher performance.

- The data taken are research publications in the last 10 years.
- The publication data processed focuses on the theme of computer network security.

– The variables used are: (percentage of similarity of publication titles with the theme of computer network security, article quartile, citation, author position, and percentage of the number of publications in the last 10 years).

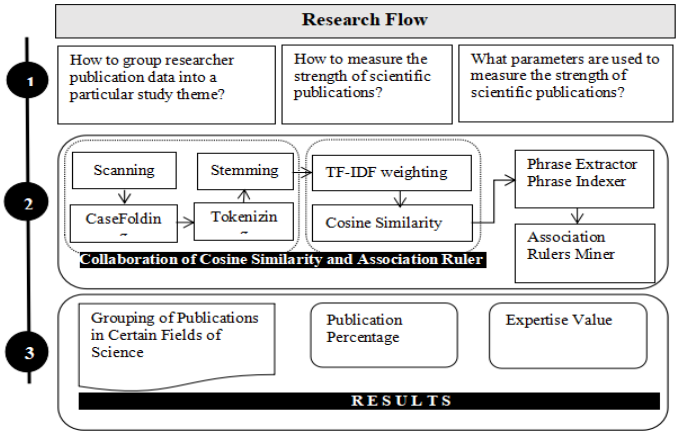

Figure 2. Collaboration flow for measuring researcher expertise

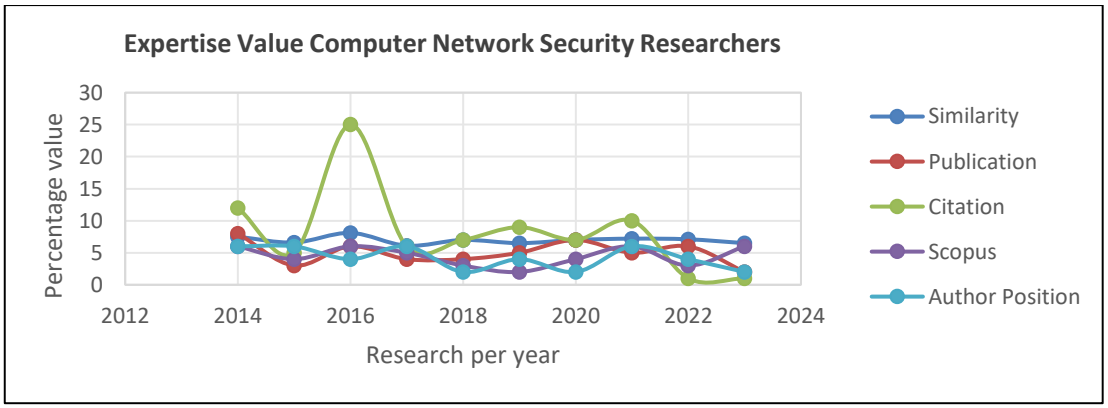

Figure 3. The percentage value of each variable of researcher publications in the last 10 years

Only research publication data is used because the research contains elements of knowledge, facts, research developments, and research recognition. In addition, each study has been tested and has been revised by the examiner. The published research results have been revised by the reviewer team. Meanwhile, data from conferences, patents, books, or other scientific works are not included, because they do not go through strict review results.

3.1. Expertise measurement

3.1.1. Similarity

The following is the implementation of the researcher's expertise measurement, starting from the detection of text similarity carried out in the study. Preprocessing is the process of changing arbitrary structured data into structured data according to needs. The preprocessing stage consists of case folding, tokenizing, filtering, stemming, tagging, and analyzing [47]. At this Processing stage, the implementation of the class is made based on data (publication title and computer network security data dictionary). Cosine similarity will calculate the level of similarity between two or more objects expressed in a vector, the number is two vectors using the keyword (cosine) [48], [49].

Cosine similarity function; cosine similarity is able to calculate text equivalence through two quantities valued in one-dimensional space cosine angle of the product of two vectors being compared, because the cosine of 0 is 1 and less than 1 for other angle values [50]. This kind of value is said to be equivalent when the cosine similarity number is 1. Cosine similarity is utilized in positive space, where the result is limited between the values 0 and 1. Cosine similarity is advantageous because even though the same

documents are not adjacent (Euclidean), they are still oriented close to each other. If the angle is small, then the similarity is high [50]. Cosine similarity overcomes the 'count-the-common-words' approach [51].

Expertise assessment and term frequency-inverse document frequency (TF-IDF) classification through the Processing stage. The class implementation is made based on the data dictionary and the level of similarity between two or more objects expressed in the vector will be calculated using as (1):

$$similarity(Q,D) = Q \cap D \vee \frac{}{Q\vee\frac{1}{2}\vee.\vee D\vee\frac{1}{2}\vee} \quad (1)$$

where, |Q∩D| is the number of words in document Q and document Y, |Q| is number of words in the data dictionary document, and |D| is number of words in the title of a published document.
Next, determining indexation/tokenization, the process calculates TF, DF, and IDF on the document.

$$idf = log(n/df) \quad (2)$$

$$(wdtweight[i] = tf[i]xidf[i] \quad (3)$$

where, *idf* is the inverse of document frequency, *df* is document frequency, *idf[i]* is IDF at index *i*, *n* is number of words per document (tf), and *tf[i]* is terms in the index *i.*

$$Q.D[i] = \sum_{j=1}^{n} wdt \quad (4)$$

$$Q \vee \sqrt{\sum_{j=1}^{n} \ wdt} \quad (5)$$

$$similarity = coscos = \frac{Q.D_i}{Q\vee D_i} \quad (6)$$

where, *wdt(Q)* is Q weight on word *j*, QD_i is vector multiplication result, $D[i]_j$ is weight D on *i* in word *j*, and $|Q||D_i|$ is calculated vector length.

3.1.2. Association rules

Association rules are a data mining technique that identifies data or text elements that frequently occur together in a data set. Association rules were first introduced by Agrawal in 1993. Next, measure the accumulation of expertise value. The task of association is useful in finding rules that are not able to cover the calculation between two or more relationships on attributes. Association rules can be described in the form: if "previous event" (then) "consequences". The calculation is also followed by support and confidence rules.

3.1.3. Implementation of expertise measurement through variables

At this stage, the accuracy of publication classification is tested using keywords (i.e., data dictionary) for each category. This test uses publication title documents and data dictionary. Cosine similarity is a measure of the angle between the document vectors Da point (ax and bx) and Db point (ay and by). Each vector represents each word in each document (text) being compared and forms a triangle [51], so that the law of cosines can be applied to state that:

$$cos(c) = a^2 + b^2 - \frac{c^2}{2ab} \quad (7)$$

where

$$a^2 = {a_x}^2 + {a_y}^2 b^2 = {b_x}^2 + {b_x}^2 \text{ and } c^2 \quad (8)$$

so that it is obtained: $\cos c = \frac{a_x b_x + a_y b_y}{\sqrt{ax^2+ay^2x}\sqrt{bx^2+by^2}}$

For two documents that are exactly, the same, the angle is zero degrees (0°) so the similarity value is one (1); and if there is no smilarity between the two, the angle is 90 degrees (90°) so similarity value is zero (0) [50].

The word list is weighted using the TF method to determine the frequency of word occurrence. The IDF of each word is calculated using (9):

$$IDF = tf_{ti} loglog\left(\frac{N}{df}\right) + 1 = 1xloglog\frac{2}{1} + 1 = 1.301 \quad (9)$$

Next, TF-IDF weighting is performed to produce weights. We will count the words “classification of darknet traffic using the AdaBoost classifier method detection”.

$$\omega d(ti) = \frac{tf\ tiloglog\left(\frac{N}{df}\right)+1}{\sum tf\ tiloglog\left(\frac{N}{df}\right)+1} = \frac{1xloglog\left(\frac{2}{1}\right)+1}{\sqrt{\left(1xloglog\left(\frac{2}{1}\right)+1\right)}} = \frac{1{,}301}{\sqrt{1x3.301}} = 07161 \quad (10)$$

The accuracy of the publication classification is tested using keywords (i.e., data dictionaries) for each category. Word weighting uses the TF method to determine the frequency of word occurrence. Then calculate the IDF using in (7). The data used is randomly taken from the official website of sinta.kemdikbud.go.id. Then measure the similarity level, where the indicator term Q is set as the category type of the data dictionary, and D is set as the title of the publication classified in the category type Q. The indicator term of category type Q is found by considering a set of terms D from all category types Q in the data. Each term is assigned a TF-IDF weight value and sorted in descending order. The top order is identified as the indicator term in the data dictionary Q.

$$coscos(Q,D) = \sum_{r=1}^{M}\ \omega_Q(t_i)x\omega_D(t_i) \quad (11)$$

$$(0.7161 * 0) + (0.5503 * 0.5503) + (0.5503 * 0.5503) + (0 * 0.7161) = 0.6056$$

After getting the similarity value of the author's publication title that meets the requirements, the next step is to accumulate the researcher's expertise value from the last ten years of publications. Starting from checking the validity of the user, then taking the last ten years of Scopus data, and matching the similarity of the publication manuscript title, then calculating the score using:

$$\left(jpt = \frac{jp}{jt}x100\%\right) + \left(q = \frac{q_1+q_2+q_3+q_4}{q}x100\%\right) + \left(o = \frac{(c+a)}{o}x100\%\right) + ct$$

where, *jpt* is number of publications per year, *q* is Scopus, *ct* is citation, *jp* is number of publications, *o* is research output, *jt* is number of the year, and *c/a* is conference/article.

In this study, data was taken directly from the SINTA Indonesia portal "https://sinta.kemdiktisaintek.go.id/," which provides access to citations and expertise in Indonesia through a web-based research information system. The data was processed to measure the TF-IDF value of a word (term) in a researcher's publication title. This value indicates how important a word is to that word in the document corpus. The results of the experiment can be seen in Table 3.

Table 3. Word frequency list TF, IDF, and TF_IDF

Author	Publication title	TF	IDF	TF-IDF $\omega_D(ti)$
1	1. The incorporation of stacked long short-term memory into intrusion detection systems for botnet attack classification [52].	1	1.301	0.7161
	2. Classification of darknet traffic using the AdaBoost classifier method detection [53].	2	1	0.5503
	3. Dimensional reduction with fast ICA for IoT botnet detection [54].	2	1	0.5503
2	1. Face recognition system using deep neural network with convolutional neural networks [55].	2	1	0.5503
	2. Investigating the ethernet and Boolean logic [56].	1	1.301	0.7161
	3. Analysis of security and performance service in service-oriented architecture (SOA) and data integration [57].	2	1	0.5503

4. RESULTS AND DISCUSSION

This study is useful for classifying publication documents based on existing categories and assessing the percentage of researcher expertise based on the dynamics of their research. This is closely related to what has been studied in depth and the researchers track record. This updated approach to measuring researcher expertise uses parameters that are directly related to the results of researcher publications. The variable parameters are; the relevance of the scientific field to research (which supports fundamental theorems to solve problems comprehensively and contextually, to have an impact on the development of science from mere information to value), the number of researcher publications according to the scientific field, publication accreditation status, researcher position (position), and the number of researcher publications each year and citations.

Figure 3 shows the results of the collaboration process of cosine similarity and association rule applied to SINTA author data by scraping by showing the results of the percentage value of the similarity of publication titles with the theme of computer network security, publications per year in the last 10 years, citations, Scopus, and author position. These variables have a significant influence on the scientific journey and expertise of researchers. The vertical line shows the percentage value of the similarity of publication titles in the field of network security per year. The horizontal line shows the year of research.

Figure 4 shows the results of the accumulation of expertise value in the field of computer network security that has been processed using collaboration of cosine similarity algorithm and associations rule. Author data is taken randomly on the SINTA web by scraping. The graph in Figure 4 shows the percentage of expertise value of each author in the last 10 years

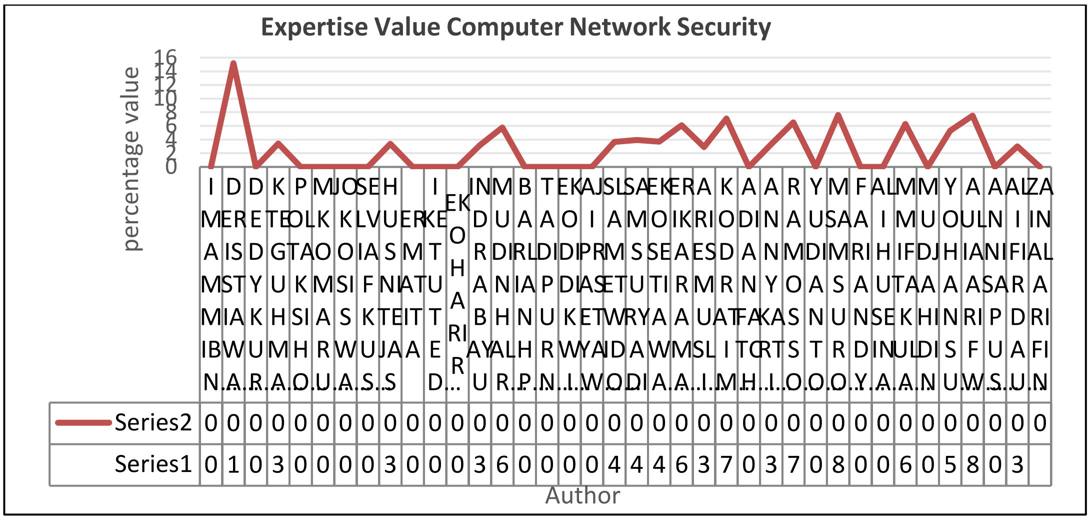

Figure 4. Results of the accumulation of expertise value using collaboration of cosine similarity and association rule

Several achievement indicators that need to be considered and have an impact on increasing international recognition include the number of citations per faculty, and international research networks. The ranking of universities in the world has become a prestigious achievement for educational institutions and countries. The achievement of a university ranking in the world's top is measured by various indicators based on ranking institutions, referring to the quality of higher education services to be able to produce human resources and innovation. The existence of world university rankings brings international quality education service standards present in many countries that contribute to the image and reputation of universities.

Figure 5 shows the influence of affiliation and support on the reputation of higher education institutions. Affiliations increase the transparency of research and provide information about the credibility of the authors and the quality of the research conducted. Good institutions have high standards in research. Affiliations are a form of recognition and support for institutions in building reputations. Affiliations also help in identifying collaborative networks between researchers that often involve cooperation between multiple institutions.

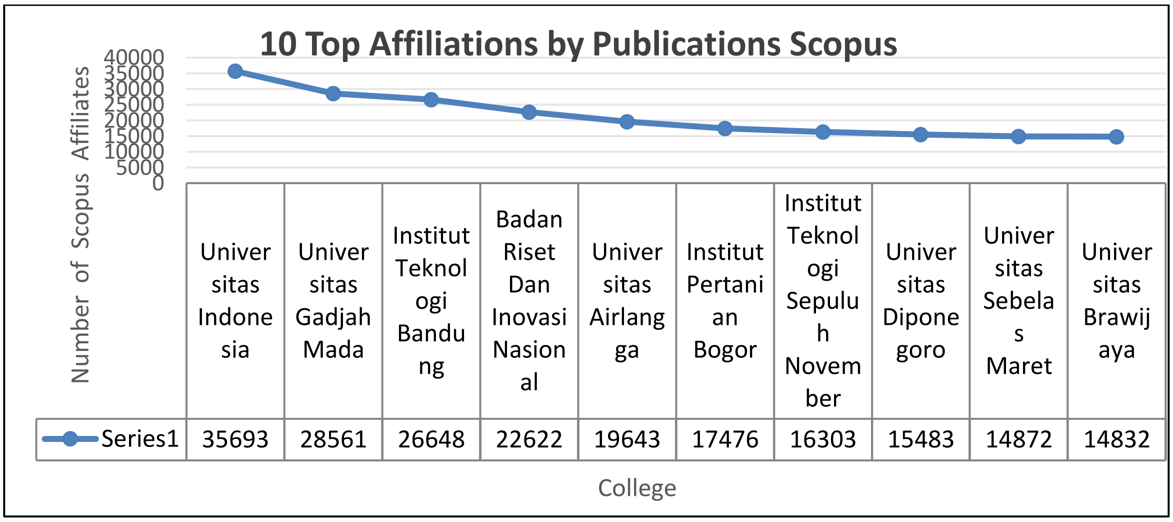

Figure 5. Affiliations by Scopus on SINTA

Affiliations serve as institutional identities, as well as providing information about the author's area of expertise and the relevance of the article to a particular research field. Clear and accurate affiliations can enhance the credibility of authors and the reputation of the institutions they are affiliated with.

5. CONCLUSION

The results of this study indicate that the research publication variables (consistency of research in the same field, article quartile, citation, author position, and percentage of the number of publications in the last 10 years), can be used as important points and parameters in measuring the level of competence and knowledge of researchers. Because publication is a form of recognition of the measurable achievements of researchers in showing the trail of scientific journey (educational background, knowledge possessed, and expertise).

In addition, the importance of the linearity of the researcher's science with his research in order to provide the latest facts so that it produces output that can be accounted for, logical processes and explanations will be easier to achieve, presentation of research based on science, producing correct data and good output. The linearity of science can eliminate inappropriate words in research such as; "maybe, could be, perhaps, if, estimate" which have the potential to eliminate or even kill what can be achieved through science, so that research becomes correct.

The accreditation status of the publication media greatly reflects the quality of the research output produced. This can be seen from the references used, results, implications, novelty of ideas, methods, and originality of the material. The position of the author in a publication is important because it has an impact on the assessment of a researcher's expertise and shows the role and contribution of the researcher to the idea, and the process until the research results are published. The main author is the owner of the idea as the basis for a study.

The routine of researcher publications will increase their expertise and knowledge. This is because every study requires a series of fundamental theorems to solve problems comprehensively and contextually which in general causes the development of science from mere information to valuable. The quality of publication reflects the output of research. An expert is required to always publish and research and must actively improve knowledge and expertise by controlling the development of science.

The image and reputation of the world's best universities "World Class University" (WCU) are closely related to the results of researchers' publications, this has become a competitiveness for universities. Scopus indexed publications affect the number of citations and h-index of researchers. Times Higher Education measures the success of universities, including using research publication indicators. The emergence of global university rankings has changed the paradigm of higher education rankings have become a standard feature to provide information and guide evaluation and planning decisions.

The quality of publications and citations is material for assessing the strength of research at institutions in measuring and comparing research performance at universities. The model built successfully validated researcher expertise up to 72% which is relevant to its support for university rankings up to 75%. Research on measuring researcher expertise still needs to be developed in all fields of scientific study, as well as seeing its relationship to the work of researchers, and the impact of the results of the research conducted.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ali Firdaus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
Deris Stiawan	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓		
Samsuryadi	✓			✓			✓			✓	✓	✓		
Rahmat Budiarto				✓		✓				✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study were taken from the website https://sinta.kemdiktisaintek.go.id and then carefully simulated to reflect the scenario of measuring reseacher expertise.

REFERENCES

[1] J. Vidal and C. Ferreira, "Universities Under Pressure: the Impact of International University Rankings," *Journal of New Approaches in Educational Research*, vol. 9, no. 2, pp. 181-193, 2020, doi: 10.7821/naer.2020.7.475

[2] A. Kelly, "The end of the line? A quantitative analysis of two decades of competitive assessment and research funding in UK higher education," *Assessment and Evaluation in Higher Education*, vol. 48, no. 5, pp. 701–722, 2023, doi: 10.1080/02602938.2022.2120961.

[3] D. Docampo, D. Egret, and L. Cram, "An anatomy of the academic ranking of world universities (Shanghai ranking)," *SN Social Sciences*, vol. 2, no. 8, pp. 1-17, 2022, doi: 10.1007/s43545-022-00443-3.

[4] N. A. Ebrahim, H. Ebrahimian, M. Mousavi, and F. Tahriri, "Does a Long Reference List Guarantee More Citations? Analysis of Malaysian Highly Cited and Review Papers," *The International Journal of Management Science and Business Administration*, vol. 1, no. 3, pp. 6–16, 2015, doi: 10.18775/ijmsba.1849-5664-5419.2014.13.1001.

[5] Y.-W. Hou and W. J. Jacob, "What contributes more to the ranking of higher education institutions? A comparison of three world university rankings," *International Education Journal: Comparative Perspectives*, vol. 16, no. 4, 29-46, 2017.

[6] E. Hazelkorn, "Reshaping the world order of higher education: the role and impact of rankings on national and global systems," *Policy Reviews in Higher Education*, vol. 2, no. 1, pp. 4–31, 2018, doi: 10.1080/23322969.2018.1424562.

[7] M. L. Germain and R. S. Grenier, "Expertise at work: Current and emerging trends," *Springer Nature Link*, 2021, doi: 10.1007/978-3-030-64371-3.

[8] A. M. Abubakar, H. Elrehail, M. A. Alatailat, and A. Elçi, "Knowledge management, decision-making style and organizational performance," *Journal of Innovation and Knowledge*, vol. 4, no. 2, pp. 104–114, 2019, doi: 10.1016/j.jik.2017.07.003.

[9] E. Santos, M. Carvalho, and S. Martins, "Sustainable Enablers of Knowledge Management Strategies in a Higher Education Institution," *Sustainability (Switzerland)*, vol. 16, no. 12, pp. 1-15, 2024, doi: 10.3390/su16125078.

[10] D. W. Aksnes, L. Langfeldt, and P. Wouters, "Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories," *SAGE Open*, vol. 9, no. 1, 2019, doi: 10.1177/2158244019829575.

[11] N. C. Brownstein, T. A. Louis, A. O'Hagan, and J. Pendergast, "The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision-Making," *American Statistician*, vol. 73, pp. 56–68, 2019, doi: 10.1080/00031305.2018.1529623.

[12] F. Joelving, "Paper Trail," *Science*, vol. 383, no. 6680, pp. 252–255, 2024, doi: 10.1126/science.ado0309.

[13] D. E. Barbosa, J. S. Paramo, and J. M. Gómez, "Academic reputation quality and research: an analysis of Latin-American universities in the world higher education institution rankings from the perspective of organizational learning theory," *Journal of Further and Higher Education*, vol. 47, no. 6, pp. 754–768, 2023, doi: 10.1080/0309877X.2023.2176204.

[14] A. I. Hermanu, D. Sari, M. C. Sondari, and M. Dimyati, "Is it necessary to evaluate university research performance instrument? Evidence from Indonesia," *Cogent Social Sciences*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2069210.

[15] P. D'Cruz, E. Noronha, and S. Katiyar, "Meaningfulness and Impact of Academic Research: Bringing the Global South to the Forefront," *Business and Society*, vol. 61, no. 4, pp. 839–844, 2022, doi: 10.1177/00076503211059761.

[16] J. Zhang, H. H. Ismail, and N. A. Sulaiman, "Exploring College English Teachers' Abilities in CLIL-Based Classrooms in Undeveloped Areas of China," *International Journal of Learning, Teaching and Educational Research*, vol. 23, no. 7, pp. 320–346, 2024, doi: 10.26803/ijlter.23.7.17.

[17] J. Emiliussen, S. Engelsen, R. Christiansen, and S. H. Klausen, "We are all in it!: Phenomenological Qualitative Research and Embeddedness," *International Journal of Qualitative Methods*, vol. 20, pp. 1-6, 2021, doi: 10.1177/1609406921995304.

[18] E. I. Fried, "Theories and Models: What They Are, What They Are for, and What They Are About," *Psychological Inquiry*, vol. 31, no. 4, pp. 336–344, 2020, doi: 10.1080/1047840X.2020.1854011.

[19] S. Hall and L. Liebenberg, "Qualitative Description as an Introductory Method to Qualitative Research for Master's-Level Students and Research Trainees," *International Journal of Qualitative Methods*, vol. 23, pp. 1-5, 2024, doi: 10.1177/16094069241242264.

[20] A. D. Vaio, R. Palladino, A. Pezzi, and D. E. Kalisz, "The role of digital innovation in knowledge management systems: A systematic literature review," *Journal of Business Research*, vol. 123, pp. 220–231, 2021, doi: 10.1016/j.jbusres.2020.09.042.

[21] T. Matthew *et al*., "Knowledge, Attitude, And Practices (KAP) Of Malaria Prevention Amongst Medical Students of Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria," *Razi International Medical Journal*, vol. 2, no. 2, pp. 28–40, 2022, doi: 10.56101/rimj.v2i2.47.

[22] A. Firdaus, D. Stiawan, and Samsuryadi, "Validity of Measurable Expertise on Publication Recognition," in *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2023, pp. 606–612, doi: 10.1109/EECSI59885.2023.10295653.

[23] M. Gusenbauer and N. R. Haddaway, "What every researcher should know about searching – clarified concepts, search advice, and an agenda to improve finding in academia," *Research Synthesis Methods*, vol. 12, no. 2, pp. 136–147, 2021, doi: 10.1002/jrsm.1457.

[24] S. Blume, "In search of experiential knowledge," *Innovation: The European Journal of Social Science Research*, vol. 30, no. 1, pp. 91–103, 2017, doi: 10.1080/13511610.2016.1210505.

[25] E. Wikström, J. Severin, I. H. Jonsdottir, and M. Akerstrom, "Process facilitators shifting between the support and expert roles in a complex work environment intervention in the Swedish healthcare sector," *Journal of Health Organization and Management*, vol. 36, no. 9, pp. 25–47, 2021, doi: 10.1108/JHOM-10-2021-0382.

[26] H. Wikström *et al.*, "Powder flow from an intermediate bulk container – Discharge predictions and experimental evaluation," *International Journal of Pharmaceutics*, vol. 597, 2021, doi: 10.1016/j.ijpharm.2021.120309.

[27] S. Marginson, "Higher education in the global knowledge economy," *Procedia - Social and Behavioral Sciences*, vol. 2, no. 5, pp. 6962–6980, 2010, doi: 10.1016/j.sbspro.2010.05.049.
[28] M. Gube and S. Lajoie, "Adaptive expertise and creative thinking: A synthetic review and implications for practice," *Thinking Skills and Creativity*, vol. 35, 2020, doi: 10.1016/j.tsc.2020.100630.
[29] J. Wilkins, B. V. Wegen, and R. D. Hoog, "Understanding and valuing knowledge assets: Overview and method," *Expert Systems with Applications*, vol. 13, no. 1, pp. 55–72, 1997, doi: 10.1016/S0957-4174(97)00022-5.
[30] C. -T. Li and M. -K. Shan, "X2-Search: Contextual Expert Search in Social Networks," in *2013 Conference on Technologies and Applications of Artificial Intelligence*, Taipei, Taiwan, 2013, pp. 176-181, doi: 10.1109/TAAI.2013.44.
[31] S. Yuan, Y. Zhang, J. Tang, W. Hall, and J. B. Cabotà, "Expert finding in community question answering: a review," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 843–874, 2020, doi: 10.1007/s10462-018-09680-6.
[32] K. Hofmann, K. Balog, T. Bogers, and M. De Rijke, "Contextual factors for finding similar experts," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 5, pp. 994–1014, May 2010, doi: 10.1002/asi.21292.
[33] V. F. He, G. V. Krogh, and C. Sirén, "Expertise Diversity, Informal Leadership Hierarchy, and Team Knowledge Creation: A study of pharmaceutical research collaborations," *Organization Studies*, vol. 43, no. 6, pp. 907–930, 2022, doi: 10.1177/01708406211026114.
[34] S. L. Ferguson, E. W. G. Moore, and D. M. Hull, "Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers," *International Journal of Behavioral Development*, vol. 44, no. 5, pp. 458–468, 2020, doi: 10.1177/0165025419881721.
[35] S. Sharma, V. Rana, and V. Kumar, "Deep learning based semantic personalized recommendation system," *International Journal of Information Management Data Insights*, vol. 1, no. 2, 2021, doi: 10.1016/j.jjimei.2021.100028.
[36] Y. D. Seo, Y. G. Kim, E. Lee, and D. K. Baik, "Personalized recommender system based on friendship strength in social network services," *Expert Systems with Applications*, vol. 69, pp. 135–148, 2017, doi: 10.1016/j.eswa.2016.10.024.
[37] L. Qiu, S. Gao, W. Cheng, and J. Guo, "Aspect-based latent factor model by integrating ratings and reviews for recommender system," *Knowledge-Based Systems*, vol. 110, pp. 233–243, 2016, doi: 10.1016/j.knosys.2016.07.033.
[38] B. Lepri, J. Staiano, E. Shmueli, F. Pianesi, and A. Pentland, "The role of personality in shaping social networks and mediating behavioral change," *User Modeling and User-Adapted Interaction*, vol. 26, no. 2–3, pp. 143–175, 2016, doi: 10.1007/s11257-016-9173-y.
[39] M. K. Najafabadi, M. N. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," *Computers in Human Behavior*, vol. 67, pp. 113–128, 2017, doi: 10.1016/j.chb.2016.11.010.
[40] D. Sharma, "Application of Association Rules In Clinical Data Mining: A Case Study For Identifying Adverse Drug Reactions," *Value in Health*, vol. 19, no. 3, p. A101, 2016, doi: 10.1016/j.jval.2016.03.1726.
[41] S. Mizzaro, M. Pavan, and I. Scagnetto, "Content-based similarity of Twitter users," *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9022, pp. 507–512, 2015, doi: 10.1007/978-3-319-16354-3_56.
[42] M. Pavan, T. Lee, and E. W. D. Luca, "Semantic enrichment for adaptive expert search," in *ACM International Conference Proceeding Series*, 2015, doi: 10.1145/2809563.2809621.
[43] H. Noureddine, I. Jarkass, H. Hazimeh, O. A. Khaled, and E. Mugellini, "CARP: Correlation-based approach for researcher profiling," in *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE*, 2015, pp. 461–464, doi: 10.18293/SEKE2015-145.
[44] C. Birtolo, D. Ronca, and G. Aurilio, "Trust-Aware Clustering Collaborative Filtering: Identification of Relevant Items," in *IFIP Advances in Information and Communication Technology*, Springer, Berlin, Heidelberg, pp. 374–384, 2012, doi: 10.1007/978-3-642-33409-2_39.
[45] A. Price *et al.*, "An accurate and practical method for assessing science and engineering problem-solving expertise," *International Journal of Science Education*, vol. 44, no. 13, pp. 2061–2084, 2022, doi: 10.1080/09500693.2022.2111668.
[46] B. S. L. Barrantes and F. J. C. Ortiz, "Quantifying the publication preferences of leading research universities," *Scientometrics*, vol. 126, no. 3, pp. 2269–2310, 2021, doi: 10.1007/s11192-020-03790-1.
[47] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020
[48] C. Amatulli, M. De Angelis, and A. Stoppani, "Analyzing online reviews in hospitality: data-driven opportunities for predicting the sharing of negative emotional content," *Current Issues in Tourism*, vol. 22, no. 15, pp. 1904–1917, 2019, doi: 10.1080/13683500.2019.1594723.
[49] R. Braun, T. Ravn, and E. Frankus, "What Constitutes expertise in research ethics and integrity?," *Research Ethics*, vol. 16, no. 1-2, pp. 1-16, 2020, doi: 10.1177/1747016119898402.
[50] M. Q. Wu, T. Y. Chen, and J. P. Fan, "Similarity measures of T-Spherical fuzzy sets based on the cosine function and their applications in pattern recognition," *IEEE Access*, vol. 8, pp. 98181–98192, 2020, doi: 10.1109/ACCESS.2020.2997131.
[51] G. Wei, J. Wang, M. Lu, J. Wu, and C. Wei, "Similarity measures of spherical fuzzy sets based on cosine function and their applications," *IEEE Access*, vol. 7, pp. 159069–159080, 2019, doi: 10.1109/ACCESS.2019.2949296.
[52] A. Heryanto *et al.*, "The incorporation of stacked long short-term memory into intrusion detection systems for botnet attack classification," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 3657–3670, 2024, doi: 10.11591/ijai.v13.i3.pp3657-3670.
[53] R. E. Sari, D. Stiawan, N. Afifah, M. Y. Idris, and R. Budiarto, "Classification of Darknet Traffic Using the AdaBoost Classifier Method," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 384–396, 2024, doi: 10.52549/ijeei.v12i2.5543.
[54] Susanto *et al.*, "Dimensional Reduction With Fast ICA for IoT Botnet Detection," *Journal of Applied Security Research*, vol. 18, no. 4, pp. 665–688, 2023, doi: 10.1080/19361610.2022.2079906.
[55] E. Fernando, D. Andwiyan, D. F. Murad, D. Touriano, and M. Irsan, "Face recognition system Using Deep Neural Network with Convolutional Neural Networks," *Journal of Physics: Conference Series*, vol. 1235, no. 1, pp. 1-8, 2019, doi: 10.1088/1742-6596/1235/1/012004.
[56] H. Rohayani, H. L. H. S. Warnars, M. Tuga, and E. Abdurahman, "Employing vector autoregression feedforward neural network with particle swarm optimization in wind speed forecasting," *Sylwan*, vol. 164, no. 3, pp. 470-480, 2020.
[57] D. Rodrigues, J. C. Estrella, and K. R. L. J. C. Branco, "Analysis of security and performance aspects in service-oriented architectures," *International Journal of Security and its Applications*, vol. 5, no. 1, pp. 13–30, 2011.

BIOGRAPHIES OF AUTHORS

Ali Firdaus Studied in Computer Engineering study program and earned a Bachelor of Computer Science from Bina Darma University in 1999, then returned to pursue computer science in the field of information systems by continuing his education until he earned a Master of Information Systems at Diponegoro University in 2011. Currently working as a Lecturer in the Computer Engineering study program, Sriwijaya State Polytechnic Palembang-Indonesia. In 2025 he completed his Doctoral program education at Sriwijaya University Faculty of Engineering, Engineering Science Study Program, Informatics Engineering Study Group. His research interests so far are in the fields of computer science, information technology and information systems, informatics technology, and multimedia. He can be contacted at email: alifirdaus@polsri.ac.id.

Deris Stiawan is currently an Associate Professor at the Faculty of Computer Science, Sriwijaya University, Indonesia. He is also a member of the IEEE and has been part of the Pervasive Computing Research Group at Universiti Teknologi Malaysia since 2010. He holds a Ph.D. in Computer Science from Universiti Teknologi Malaysia and an M.Eng. in Electrical Engineering from Universitas Gadjah Mada. In addition, he is a Certified Ethical Hacker (C|EH) and Certified Hacker Forensic Investigator (C|HFI) licensed by EC-Council USA, and a Cisco Certified Networking Associate since 2005. His research interests include computer networks, intrusion detection/prevention systems, and heterogeneous networks. His research focuses on computer and network security, especially network attack and intrusion detection/prevention systems. He can be contacted at email: deris@unsri.ac.id.

Samsuryadi is an Associate Professor in the Department of Informatics Engineering at Sriwijaya University, Indonesia. He earned his Bachelor of Science S.Sc. from Sriwijaya University in 1995, Master of Computer Science M.Kom. from Gadjah Mada University in 2002, and Ph.D. in Computer Science from Universiti Teknologi Malaysia in 2013. His research interests include image processing, machine learning, and pattern recognition. He has contributed to various studies, including the development of an efficient MobileNet architecture for image recognition on mobile and embedded devices. He can be contacted at email: syamsuryadi@unsri.ac.id.

Rahmat Budiarto received B.Sc. degree in Mathematics from Bandung Institute of Technology, Indonesia in 1986, M.Eng. and Dr.Eng. in Computer Science from Nagoya Institute of Technology, Japan in 1995 and 1998, respectively. Currently, he is a full professor at Department of Computer Science, College of Computer and Information, Albaha University, KSA. His research interests include intelligent systems, brain modeling, IPv6, network security, wireless sensor networks, and MANETs. He can be contacted at email: rahmat@bu.edu.sa.