TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 23, No. 4, August 2025, pp. 954~964
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i4.26221

Business intelligence system model to measure the performance of lecturers' scientific publications

Miwan Kurniawan Hidayat[1,2], Dedy Sugiarto[2,3], Rina Fitriana[2,4], Yun-Chia Liang[5]
[1]Study Program of Industrial Engineering, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Jakarta, Indonesia
[2]Master of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jakarta, Indonesia
[3]Study Program of Information Systems, Faculty of Industrial Technology, Universitas Trisakti, Jakarta, Indonesia
[4]Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jakarta, Indonesia
[5]Department of Industrial Engineering and Management, College of Engineering, Yuan Ze University, Taoyuan City, Taiwan

Article Info

Article history:

Received Mar 24, 2024
Revised Apr 8, 2025
Accepted May 10, 2025

Keywords:

Business intelligence
Data mining
Data warehouse
K-means
X-means

ABSTRACT

Scientific publication data is sourced from the SINTA website to measure the performance of journals, institutions, and researchers in Indonesia. Accessing and analyzing data for institutional needs is restricted, and lecturer development patterns based on lecturer characteristics remain untapped. The study aims to analyze and design business intelligence system models to measure the performance of scientific publications using dimensional models, clustering, on-line analytical processing (OLAP), and prototyping. Research methods are carried out through data and information needs analysis, data warehouse design, data mining and OLAP application, business intelligence system development, and system evaluation. The resulting dimensional models are the researcher index model, the researcher score model, the publication article model, and the research subject model. Measurements of data size and processing time show that the star schema has data of 336 KB and a processing time of 0.00554 seconds, is the best model compared to the snowflake's schema, which has data of 368 KB and a processing time of 0.00611 seconds. Davies–Bouldin Index (DBI) measurements show the best clustering performance is the X-means algorithm with K as many as 5 clusters (Kmin=3, Kmax=5) and a DBI value of 0.537040.

Corresponding Author:

Rina Fitriana
Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti
St. Kyai Tapa No.1, Grogol, Jakarta, Indonesia
Email: rinaf@trisakti.ac.id

1. INTRODUCTION

As a higher education institution, the university strives to enhance lecturer resources by assessing their research output in scientific publications to improve education quality. Scientific publications serve as the final stage in disseminating research findings, contributing to problem-solving where no prior solution exists. According to Indonesian Law No. 14 of 2015, lecturers are required to conduct scientific publications, with universities setting indicators for research output standards, including publications in accredited national and reputable international journals, citation impact, and alignment with research roadmaps. University leaders measure publication performance based on data from indexing institutions such as the Science and Technology Index (SINTA), Scopus, Web of Science, the Directory of Open Access Journals (DOAJ), ProQuest, the ASEAN Citation Index (ACI), Google Scholar, and the Indonesian Publication Index (IPI). The SINTA platform, managed by the Ministry of Education, provides research performance insights for researchers, institutions, and journals. However, as SINTA operates as an external system, universities require an

independent data storage system tailored to internal needs, ensuring better access and analysis for monitoring lecturers' scientific publications. The H-index is a bibliometric indicator widely used in academia, research management, and policy. In simple terms, the H-index combines measures of quantity and impact in one indicator [1]. The problem in this research is the limited access and analysis of data from the SINTA website, which is managed by external systems, making it difficult for higher education institutions to meet their internal information needs for evaluating the performance of academic publications by lecturers. As a means to resolve this issue, it is suggested to create a business intelligence system model that assesses the performance of lecturers' scientific publications using a dimensional model along with data mining methods, thereby allowing higher education institutions to efficiently access, analyze, and leverage data to satisfy their internal information needs. In designing the business intelligence and data mining system model to monitor faculty scientific publications, there are challenges in determining the best dimensional model between star schema and snowflake schema, as well as the most optimal clustering algorithm between K-Means and X-means.

Previous studies have explored different aspects of business intelligence systems, though with certain constraints. Some focused on these systems without comparing dimensional models and data mining algorithms, while others utilized a single data mining algorithm without examining dimensional model alternatives. There have also been comparisons of dimensional models without involving data mining, and research comparing data mining algorithms without incorporating them into a business intelligence system. This research presents a fresh approach by more comprehensively integrating business intelligence systems with data mining algorithms. It not only applies data mining techniques but also compares K-means and X-means to determine the most effective algorithm. Additionally, the study contrasts two major dimensional models, star schema and snowflakes schema, to identify the best model for evaluating lecturers' scientific publication performance. The selection of methods in research refers to previous studies that are tailored to the needs and conditions of the organization, namely comparative research on dimensional model performance [2]; an optimized hybrid approach for enhancing the extract-transform-load (ETL) process efficiency in a data warehouse [3]; X-means clustering research using Davies–Bouldin Index (DBI) evaluation [4]; quality monitoring in higher education with business intelligence dashboards [5]; star-schema research model for lecturer performance in research activities [6]; K-means clustering approach for analyzing customer electricity consumption patterns [7]; a study discussing the integration of data mining within business intelligence to enhance understanding of customer needs [8].

This study focuses on designing an effective business intelligence and data mining system model to monitor faculty scientific publications. The research has two main objectives, which are to determine the best dimensional model between star schema and snowflake schema for managing publication data and to identify the most optimal clustering algorithm between K-means and X-means for accurately and efficiently grouping publication data. Business intelligence is essential for companies of all sizes, providing valuable insights that are key to making timely decisions [9]. To remain competitive and achieve success, businesses must develop analytical strategies that transform these challenges into effective business systems [10]. Business intelligence is a term that merges architectures, methodologies, applications, analysis tools, databases, and tools to access data interactively, enable data manipulation, and for analysts and business managers to perform appropriate analysis [11]. Some of the benefits of using business intelligence are that it helps to view reports independently; helps to identify waste in the system; helps to identify the strengths and weaknesses of the company; improves decision-making processes; enables real-time analysis with quick navigation; facilitates sharing and access of information; provide very fast answers to business questions and problems; provides data visualization so that it is very easy to read, understand, and interpret [12].

In many industries, business intelligence is now widely regarded as a key driver in understanding organizational outputs well and measuring them in real-time to make improvements and changes. Business intelligence as a support in operational and strategic decision-making that can be connected to other information systems, namely customer relationship management (CRM) marketing, visualization, decision support systems (DSS)/executive information systems (SIS), data warehouse, knowledge management, data mining, on-line analytical processing (OLAP), geographic information systems (GIS) [13]. Business intelligence systems have four main components, namely user interface (e.g., data visualization); business performance management (BPM); business analytics; and data warehouse [11]. A data warehouse is unified data from various origins for dimensional data analysis. More technically, a data warehouse is a data collection with a subject, integrated, non-volatile approach with time variations to support management decisions [14]. The relational design of the data warehouse is represented in a dimensional model consisting of a star schema and a snowflake schema. The star schema consists of one centralized fact table and dimension tables. In general, dimension tables in the star schema are not normalized and may contain redundant data. Snowflake schema avoids redundancy in the star schema through dimensional representations of multiple tables associated with referential integrity constraints [14].

Smaller data warehouse sizes and shorter query processing times can determine better dimensional model performance. Small data warehouse size leads to less memory consumption [2]. Data visualization uses

visual representations to explore, understand, and communicate data. Although the name commonly used is data visualization, it means information visualization. Information refers to the contextualized, summarized, and aggregated form of raw data; thus, what appears in a visualization is information rather than raw data itself. Data visualization is inherently linked to statistical charts, scientific visuals, and information graphics [11]. Data mining is defined as the automated process of extracting meaningful insights from large-scale data repositories [15]. This process emphasizes uncovering hidden patterns and relationships to support informed decision-making [16]. The implementation of machine learning in data mining has been widely adopted across industries such as finance, retail, insurance, and telecommunications [17]. Machine learning in this context is categorized into four types: reinforcement, semi-supervised, supervised, and unsupervised learning [18]. Data mining techniques are effective tools to group agricultural product customers based on their purchase behavior [19]. Clustering algorithms are data mining methods used to identify inherent patterns within data and to provide a simplified representation of datasets [20]. K-means is a clustering algorithm that organizes data points into groups by minimizing the distance between each data point and the centroid of its cluster [21].

The K-means algorithm begins by selecting the number of clusters (K), initializing centroids, assigning data to the nearest centroid, and iteratively updating the centroids. In contrast, X-means extends K-means by automatically determining the optimal number of clusters based on the bayesian information criterion (BIC). This method requires setting a minimum (Kmin) and maximum (Kmax) number of clusters, from which it identifies the best K value. X-means operates in two main stages: Improve-Params, which runs K-means until convergence, and Improve-Structure, which evaluates whether clusters should be split using BIC. Validating clustering results is crucial for ensuring reliable application outcomes. Various internal validation metrics are available, totaling at least twelve, including coefficient of variation nearest neighbor (CVNN), scattering-density between and within clusters (S_Dbw), scatter density (SD), Xie–Beni index (XB), DBI, Silhouette coefficient (S), Dunn index (D), I index, Calinski–Harabasz index (CH), Gamma index (Γ), R-squared (RS), and root mean square standard deviation (RMSSTD) [22]. Among them, the DBI is commonly used to assess cluster quality. Cohesion in this context refers to how closely data points are grouped around the center of their cluster, and optimal clustering is indicated by a DBI value approaching zero [4].

This study aims to design an efficient business intelligence and data mining system model to track faculty scientific publications, ensuring effective data organization and insightful analysis. It explores the selection of the most appropriate dimensional model, evaluating the advantages of star schema and snowflake schema in optimizing data structuring, retrieval speed, and query performance. Additionally, the study seeks to determine the most effective clustering algorithm, comparing K-means and X-means based on their accuracy and computational efficiency in categorizing publication data. The research methodology involves a comparative evaluation of both dimensional models using performance metrics, while clustering techniques are assessed through validity indices and efficiency analysis. The expected outcome is a scalable and reliable system that enhances data-driven decision-making, enables comprehensive faculty publication monitoring, and strengthens research assessment processes within academic institutions.

2. METHOD

This section explains how about the research steps, the data collection and data warehouse design phase in this research refers to the data warehouse process stages used in the study of massively parallel processing systems to enhance the scalability of data warehouse processing [23]. The selection of the best dimensional model between star schema and snowflake schema refers to research on the advantages and disadvantages of both models [2]. OLAP technology supports the design and examination of complex multidimensional structures within business data [24]. The business intelligence system integration phase is adapted from the methodology in the study of business intelligence solutions for online banking companies [25], while the system evaluation phase uses the system usability scale (SUS) method based on the usability evaluation of e-Government websites [26]. Figure 1 explains the flow of stages in the research conducted.

2.1. Data and information needs analysis

Qualitative descriptive is used to provide simple and clear explanations of phenomena [27]. This approach helps explain how business intelligence systems function and support organizational goals. In the data needs analysis stage, data is identified and collected through observation and document review related to lecturers' scientific publications. Secondary data is obtained through data extraction from the university system and the SINTA website [28]. Web scraping is used to automate data collection into structured formats [29], making research more efficient [30].

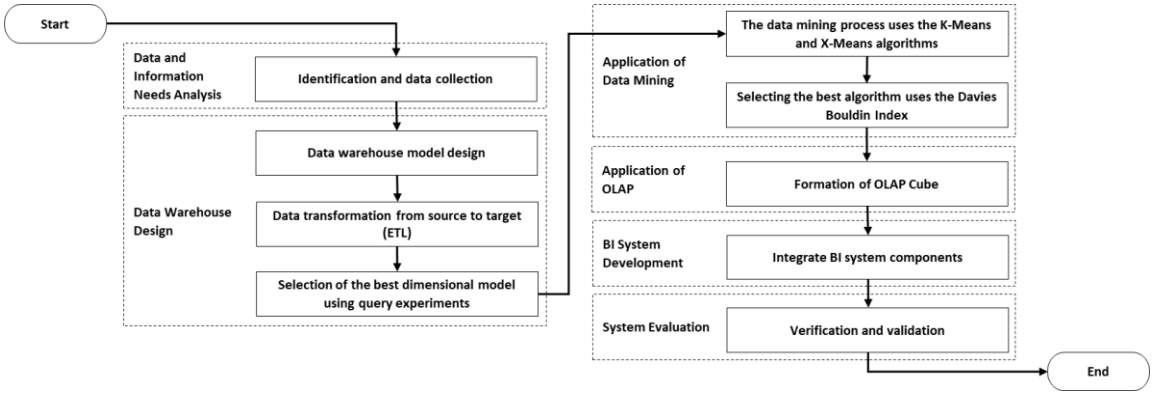


Figure 1. Stages in the research

2.2. Data warehouse design

This phase includes several steps. The data warehouse model design applies two-dimensional models: the star schema, consisting of fact and dimension tables, and the snowflake schema, which adds sub-dimension tables; both are implemented in a database. Data transformation is then carried out from source to target through ETL processes using specific software. The best dimensional model is selected based on experimental results, where smaller data sizes and faster query times indicate better performance.

2.3. Application of data mining

The application of data mining through the clustering process of lecturers as researchers involves several stages. First, clustering is performed using K-means and X-means algorithms based on H-index values from Scopus, Google Scholar, and Web of Science on the SINTA website, utilizing data mining software. Then, the best clustering algorithm is selected by comparing the DBI values, which serve as a measure of cluster validity in clustering methods.

2.4. Application of OLAP

OLAP cubes facilitate the visualization, exploration, and navigation of data within data warehouses to enhance business intelligence and forecasting [31]. Alongside business intelligence systems, OLAP supports multi-perspective analysis through various queries and operations [32]. Its core structure, the OLAP cube, is a multidimensional model that enables rapid data analysis [11]. OLAP cubes represent data to fulfill informational needs in decision-making processes.

2.5. Business intelligence system development

The development of a business intelligence system involves the integration of several key components. A data warehouse functions as the central data source, which is processed by business analytics tools to generate meaningful information. This information is then presented through a user interface using data visualizations in the form of dashboards, in accordance with the architecture of the designed business intelligence system.

2.6. System evaluation

Verification and validation are essential activities in evaluating business intelligence systems. Verification ensures the system runs without errors and meets predefined specifications by comparing test outcomes with success indicators. A system passes verification if it functions properly without defects. The SUS is used as a standard to evaluate usability, offering quantitative metrics for statistically analyzing user experience [33].

3. RESULTS AND DISCUSSION

Business Intelligence serves as a key tool for organizations in managing open innovation by enabling data analysis that generates insights and fosters innovation [34]. The business intelligence system developed in this study consists of integrated components designed to support decision-making through the monitoring of scientific publication performance. At the core of this system is a data warehouse that functions as a centralized data source, where raw data is processed using business analytics tools. These tools transform the data into meaningful information that reflects the research performance of faculty members and study

programs. The output of this processing is presented through a user-friendly interface in the form of interactive dashboards. These dashboards are aligned with the principles of business performance management (BPM) and serve to visualize trends, patterns, and key performance indicators related to scientific publications. The visual presentation not only enhances accessibility and comprehension for decision-makers but also supports strategic planning in academic institutions.

In Figure 2, inputs in the scientific publication business intelligence system (SiBIPI) include lecturer and study program data, researcher indexation, and publication article data. The absence of business intelligence and data warehouse systems presents physical barriers. Stakeholders such as research institute officers and heads of study programs act as system users. The system aligns with the university's vision, mission, research roadmap, and government regulations. SiBIPI aims to improve efficiency, effectiveness, and optimization, producing outputs like researcher indexes, scores, and publication data, resulting in a functional prototype. The designed BI system architecture model is illustrated in Figure 3.

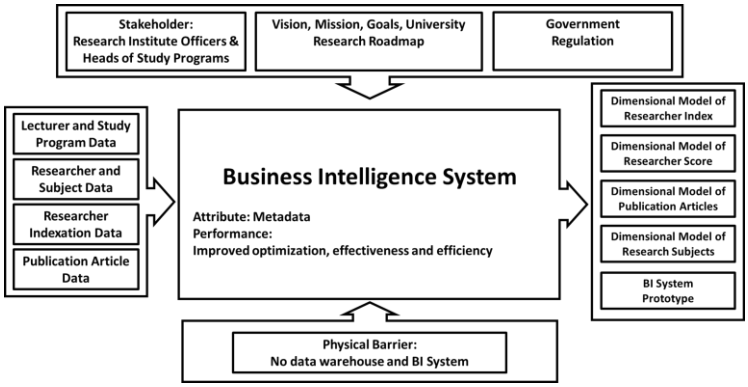


Figure 2. Business intelligence system

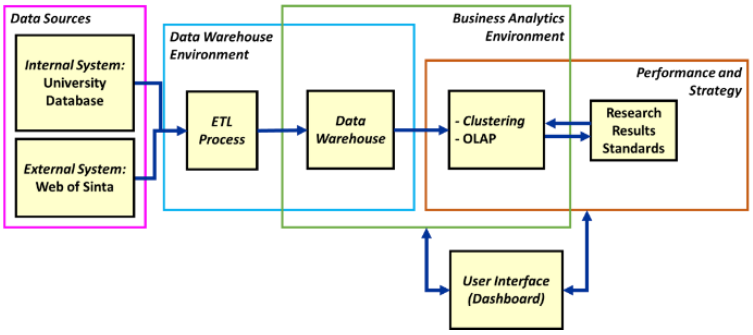


Figure 3. Business intelligence system architecture model

3.1. Data and information needs analysis

After identifying the required data and information for designing the business intelligence system, data collection is conducted on lecturers' scientific publications sourced from organizational documents, university databases, and the SINTA website. The data used as a source in the data warehouse includes lecturer and study program data, researcher data and research subjects, researcher indexation data, and publication article data. This data serves as the foundation for building an effective and structured business intelligence system that supports institutional goals. Data collection is conducted through extraction from the university database and the SINTA website using web scraping techniques. Data storage in a data warehouse built on MySQL can offer effective performance [35]. Web scraping is automated using Python programs, which are widely used for data processing and extraction tasks [36]. The information needed includes researcher index data, researcher scores, publication article data, and research subject data, all categorized by lecturers, study programs, indexers, years, and research roadmaps.

3.2. Data warehouse design

Data warehouse projects use the ETL method with a predefined schema before storing data, known as "schema-on-write" [37]. This study's data warehouse design includes model design, data transformation ETL, and model selection. A centralized architecture is applied, with data sourced from databases and spreadsheet files. The ETL process extracts, transforms, and loads data into the warehouse, which users can access through an application interface.

The data model design uses two-dimensional models: the star schema and the snowflake schema. The dimensional model design process consists of four steps: selecting business processes, establishing grain, identifying dimensions, and identifying facts [38]. In the scientific publication monitoring process, grain and dimension identification results in a relationship used to form fact tables. These fact tables are constructed based on the grain and its related dimensions to support data analysis.

One of the dimensional models designed is the researchers' score model, which presents researcher scores based on lecturers, study programs, and clustering through relationships between score facts and related dimensions. Figure 4(a) show the dimensional score model researchers use the star schema, and Figure 4(b) show the dimensional score model researchers use the snowflakes schema.

One of the dimensional models designed is the researchers' score model, which presents researcher scores based on lecturers, study programs, and clustering through relationships between score facts and related dimensions. Figure 4(a) show the Dimensional Score model researchers use the Star Schema, and Figure 4(b) show the Dimensional Score model researchers use the Snowflakes Schema.

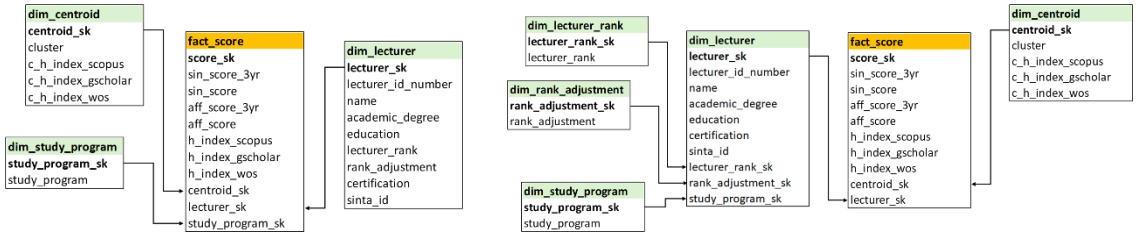


Figure 4. Dimensional model of researcher score: (a) star schema and (b) snowflakes schema

The ETL process is essential for loading data into the data warehouse, ensuring it meets defined standards. Various ETL tools support data extraction from different sources and manage initial, historical, and incremental loading, each with unique structures and functions [39]. In this study, ETL processes for dimension and fact tables in star and snowflake schema models are handled using Pentaho Data Integration. For instance, the ETL process in the article fact table uses components like table input, add sequence, select values, and table output.

Various organizations have utilized MySQL to develop extensive database systems [40]. MySQL software is used as a database management system in a data warehouse. Data on star schema models and snowflakes schema models are stored in separate databases. Experiments on the same data and results from applying queries were performed on both dimensional models to find out which model performed better between the star schema and snowflakes schema in the data warehouse for this case. The query experiments consist of 4 queries, including information about the index, the score, the published article, and the subject of research. The total query processing time for the star schema model is 0.00554 seconds, while the total query processing time for the snowflakes schema model is 0.00611 seconds. A comparison of experimental results can be seen in Table 1. The comparison between the star schema model and the snowflakes schema in the experiment shows that the star schema model is the best model chosen because it has a smaller data size of 336 KB and a shorter query processing time of 0.00554 seconds.

Table 1. Comparison of dimensional models

Model dimensional	Total size (KB)	Runtime (sec)
Star schema	336	0.00554
Snowflakes schema	368	0.00611

3.3. Application of data mining

The application of data mining aims to group researchers based on H-index values in Scopus, Google Scholar, and Web of Science. Researchers are grouping by applying two clustering algorithms, namely the

K-means and X-means algorithms. The data used for clustering is sourced from a data warehouse that has undergone a transformation process and has been cleaned. K-means and X-means algorithms are used with clustering (K) as many as 5 clusters at the clustering modeling stage. Grouping is done gradually with K values ranging from 2, 3, 4, and 5. The clustering process uses data mining software. Python programs were used as data mining software in this study.

Cluster validation using the DBI method on the K-means and X-means algorithms with a grouping (K) of 5 clusters is calculated during the clustering process using Python programs as data mining software. Cluster validation results are compared to select the best clustering. Table 2 compares DBI values after the clustering process. The results of the comparison of DBI values show that the best clustering performance is the X-means clustering algorithm with optimal grouping (K) of 5 clusters (Kmin=3, Kmax=5) and DBI value of 0.537040.

Table 2. DBI value comparison

K-means clustering		X-means clustering		
K/cluster	DBI value	K	Cluster	DBI value
K=2	0.778144	Kmin=2, Kmax=5	4	0.614357
K=3	0.682411	Kmin=3, Kmax=5	5	**0.537040**
K=4	0.655393	Kmin=4, Kmax=5	5	0.553758
K=5	0.734986	Kmin=5, Kmax=5	5	0.747751

3.4. Application of OLAP

An OLAP cube was developed as a data analysis technique in this business intelligence system to fulfill functional requirements in delivering essential information. The system includes several OLAP cubes, such as the researcher index cube, researcher score cube, publication articles cube, and research subject cube. Specifically, the publication articles cube provides detailed information about published articles, as illustrated in Figure 5.

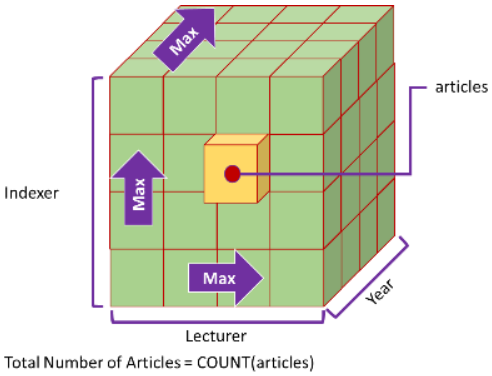


Figure 5. Publication articles cube

3.5. Business intelligence system development

The business intelligence system design includes components that generate information to measure scientific publication performance through data visualization dashboards. Power BI is used to build the system, enabling detailed data processing and interactive graphical visualization. One example is the researcher score dashboard, which contains various data visualization elements. The display of the researcher score page in the developed system is shown in Figure 6.

3.6. System evaluation

System evaluation is carried out to ensure the system is free from errors and can produce outputs according to user expectations and needs. Verification testing confirms that the system operates based on predetermined specifications and contains no defects. The test results show that the business intelligence system was successfully verified. Validation is conducted using the SUS method to assess user satisfaction. With a SUS score of 86.5, the system falls into the excellent category and is deemed suitable for use.

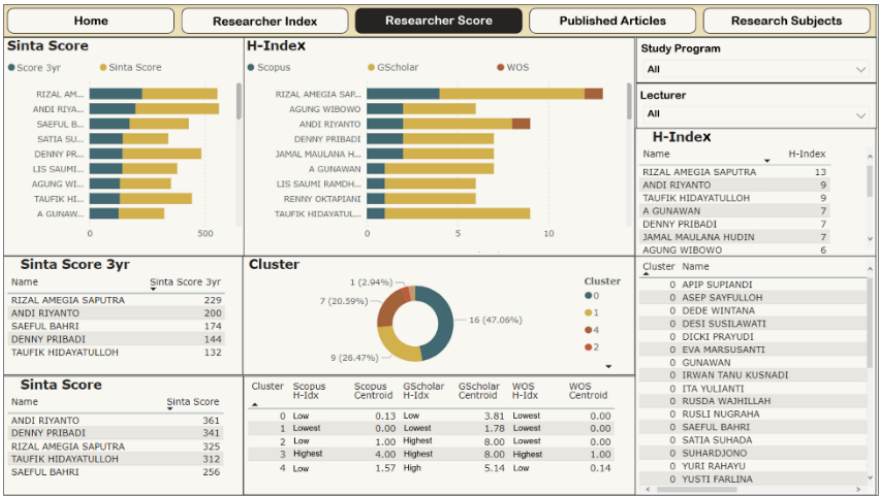


Figure 6. Researcher score page view

3.7. Comparison with related studies

This study sets itself apart by evaluating both dimensional models (star schema and snowflake schema) and clustering algorithms (K-means and X-means), unlike many previous studies that focused only on one aspect. Prior research, such as [2], [6], did not comprehensively address both areas. The main strength of this study lies in its integrated approach, assessing data models, data visualization, and clustering performance. However, this research does not explore alternative clustering algorithms such as self-organizing map (SOM), DBSCAN, or K-Medoids. While X-means outperformed K-means, the improvement in clustering performance was relatively modest, suggesting that further investigation into other clustering techniques could be valuable.

3.8. Limitations and future work

This study focused on developing a business intelligence and data mining system model to monitor faculty scientific publications, with an emphasis on identifying the most effective dimensional model and the best clustering algorithm. The results indicated that the star schema was more efficient for handling publication data, offering smaller storage requirements and faster query times, while the X-means algorithm outperformed K-means in terms of clustering accuracy. The findings underscore the importance of this research for academic institutions, as it provides a more streamlined approach to managing and analyzing publication data, aiding in the evaluation of faculty performance. Nonetheless, further exploration is needed to assess how these models perform with larger datasets and more complex computing environments. Future studies could investigate additional factors such as memory and CPU usage, as well as the application of more sophisticated machine learning algorithms to enhance decision-making and generate deeper insights.

4. CONCLUSION

This study successfully designed a business intelligence and data mining system model to monitor faculty scientific publications by selecting the most efficient dimensional model and the optimal clustering algorithm. The findings confirm that the star schema is the superior choice for structuring publication data, as it minimizes data size (336 KB) and enhances query speed (0.00554 seconds), making it more efficient than the snowflake schema. Meanwhile, the X-means clustering algorithm outperforms K-means, achieving a lower DBI (0.537040) and optimal clustering into 5 groups, demonstrating its superiority in accurately classifying faculty publication data. The adoption of star schema and X-means clustering provides higher education institutions with a more efficient data processing framework, enabling improved faculty performance evaluation and data-driven decision-making. This system enhances the analysis of scientific publications, facilitating better academic resource management and institutional research strategies. Future research should explore scalability with larger datasets, incorporating additional performance metrics such as memory usage, CPU consumption, and network load to further validate the best dimensional model for large-scale data warehouses. Moreover, integrating advanced analytics and machine learning techniques could provide deeper insights, improving strategic decision-making in academic institutions.

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
Miwan Kurniawan Hidayat	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓
Dedy Sugiarto	✓	✓		✓	✓	✓		✓	✓	✓				✓
Rina Fitriana	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓
Yun-Chia Liang	✓	✓		✓	✓	✓		✓	✓	✓				

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 obtained from a third party under a special license. Data are available from the authors with the permission of the third party. Access to these data is restricted and can be made available upon reasonable request with the permission of the third party.

REFERENCES

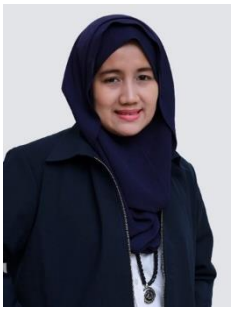
[1] G. Fraumann and R. Mutz, *The h-index*. Handbook Bibliometrics | Edition: De Gruyter Reference, 2021.
[2] M. Z. Iqbal, G. Mustafa, N. Sarwar, S. H. Wajid, J. Nasir, and S. Siddque, "A review of star schema and snowflakes schema," in *Intelligent Technologies and Applications (INTAP 2019)*, 2020, pp. 129–140. doi: 10.1007/978-981-15-5232-8_12.
[3] L. Dinesh and K. G. Devi, "An efficient hybrid optimization of ETL process in data warehouse of cloud architecture," *Journal of Cloud Computing*, vol. 13, no. 1, Jan. 2024, doi: 10.1186/s13677-023-00571-y.
[4] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering X-means algorithm with davies-bouldin index evaluation," *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, Jan. 2020, doi: 10.1088/1757-899X/725/1/012128.
[5] A. Sorour and A. S. Atkins, "Big data challenge for monitoring quality in higher education institutions using business intelligence dashboards," *Journal of Electronic Science and Technology*, vol. 22, no. 1, Mar. 2024, doi: 10.1016/j.jnlest.2024.100233.
[6] M. M. Amin, A. Sutrisman, and Y. Dwitayanti, "Development of star-schema model for lecturer performance in research activities," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 74–80, 2021, doi: 10.14569/IJACSA.2021.0120909.
[7] Y. Chen, P. Tan, M. Li, H. Yin, and R. Tang, "K-means clustering method based on nearest-neighbor density matrix for customer electricity behavior analysis," *International Journal of Electrical Power & Energy Systems*, vol. 161, Oct. 2024, doi: 10.1016/j.ijepes.2024.110165.
[8] M. A. Khder and I. A. Abu-AlSondos, "Business Intelligence and Data Mining: Opportunities and Future," *European Journal of Business and Management*, vol 13, no 11, 2021, doi: 10.7176/EJBM/13-11-01.
[9] R R. Subramaniam, P. Palakeel, M. Arunmozhi, M. Sridharan, and U. Marimuthu, "Factors driving business intelligence adoption: an extended technology-organization-environment framework," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 3, pp. 1893–1903, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp1893-1903.
[10] A. E. Nurhidayat, R. Fitriana, D. Suhardini, and A. Nugrahanti, "The causal loop diagram model of traceability system rental equipment in oil and gas supporting companies," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 5, pp. 1104–1112, Jul. 2024, doi: 10.12928/telkomnika.v22i5.26084.
[11] R. Sharda and E. Turban, *Business intelligence, analytics, and data science: A managerial perspective*. Pearson Education Limited, 2018.
[12] M. Joshi and A. Dubbewar, "Review on business intelligence, its tools and techniques, and advantages and disadvantages," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 12, pp. 386–391, 2021.
[13] M. Ur Rehman *et al.*, "Elaborating the role of business intelligence (BI) in healthcare management," *Journal of Intelligence Studies in Business*, vol. 12, no. 2, pp. 26–35, Feb. 2023, doi: 10.37380/jisib.v12i2.952.
[14] A. Vaisman and E. Zimányi, *Data warehouse systems: Design and implementation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-642-54655-6.

[15] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to data mining second edition*. Pearson Education Limited, 2019.
[16] R. A. Ramadhani, R. Fitriana, A. N. Habyba, and Y. -C. Liang, "Enhancing quality control of packaging product: A six sigma and data mining approach," *Jurnal Optimasi Sistem Industri*, vol. 22, no. 2, pp. 197–214, Dec. 2023, doi: 10.25077/josi.v22.n2.p197-214.2023.
[17] X. Teng and Y. Gong, "Research on application of machine learning in data mining," *IOP Conference Series: Materials Science and Engineering*, vol. 392, no. 6, Aug. 2018, doi: 10.1088/1757-899X/392/6/062202.
[18] M. M. Zaben, M. Y. Worku, M. A. Hassan, and M. A. Abido, "Machine learning methods for fault diagnosis in AC microgrids: A systematic review," *IEEE Access*, vol. 12, pp. 20260–20298, 2024, doi: 10.1109/ACCESS.2024.3360330.
[19] T.-C. Chen *et al.*, "Application of Data Mining Methods in Grouping Agricultural Product Customers," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–9, Mar. 2022, doi: 10.1155/2022/3942374.
[20] H. Xu, "Research on clustering algorithms in data mining," in *2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, IEEE, Jul. 2022, pp. 652–655, doi: 10.1109/ICBAIE56435.2022.9985831.
[21] A. Poggiali, A. Berti, A. Bernasconi, G. M. Del Corso, and R. Guidotti, "Quantum clustering with K-means: A hybrid approach," *Theoretical Computer Science*, vol. 992, Apr. 2024, doi: 10.1016/j.tcs.2024.114466.
[22] H. Xiong and Z. Li, "Clustering validation measures," in *Data Clustering*, Chapman and Hall/CRC, 2018, pp. 571–606. doi: 10.1201/9781315373515-23.
[23] F. C. D. Bani, Suharjito, Diana, and A. S. Girsang, "Implementation of database massively parallel processing system to build scalability on process data warehouse," *Procedia Computer Science*, vol. 135, pp. 68–79, 2018, doi: 10.1016/j.procs.2018.08.151.
[24] V. A. Sokolov, R. I. Kuzmich, A. A. Stupina, K. A. Ponomareva, and M. V. Pokushko, "OLAP Concept as a Basis for Business Analysis of Multidimensional Data Structures," *European Proceedings of Computers and Technology*, Feb. 2023, pp. 1–10, doi: 10.15405/epct.23021.1.
[25] A. E. Viteri, J. G. Cruzado, and L. A. Huaman, "Methodology for business intelligence solutions in internet banking companies," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 12, no. 3, pp. 1173–1181, May 2022, doi: 10.18517/ijaseit.12.3.13670.
[26] A. Ilyas, S. H. Wajid, and A. Muhammad, "Usability evaluation of E-government website: A use of system usability scale," *Pakistan Journal of Engineering and Technology*, vol. 5, no. 1, pp. 11–15, Mar. 2022, doi: 10.51846/vol5iss1pp11-15.
[27] A. Tunsi *et al.*, "Family caregivers' experiences of caring for patients with heart failure: A qualitative study," *Journal of the Practice of Cardiovascular Sciences*, vol. 10, no. 1, pp. 9–17, Jan. 2024, doi: 10.4103/jpcs.jpcs_54_23.
[28] F. V. Fernandez, "Intelligent information extraction from scholarly document databases," *Journal of Intelligence Studies in Business*, vol. 10, no. 2, pp. 44–61, Jun. 2020, doi: 10.37380/jisib.v10i2.584.
[29] S. Goulas and G. Karamitros, "How to harness the power of web scraping for medical and surgical research: An application in estimating international collaboration," *World Journal of Surgery*, vol. 48, no. 6, pp. 1297–1300, Jun. 2024, doi: 10.1002/wjs.12220.
[30] B. C. von Hohenberg, S. Stier, A. S. Cardenal, A. M. Guess, E. Menchen-Trevino, and M. Wojcieszak, "Analysis of web browsing data: A guide," *Social Science Computer Review*, vol. 42, no. 6, pp. 1479–1504, Dec. 2024, doi: 10.1177/08944393241227868.
[31] S. Mathur, S. L. Gupta, and P. Pahwa, "Optimizing OLAP cube for supporting business intelligence and forecasting in banking sector," *Journal of Information Technology Management*, vol. 13, no. 1, pp. 81–99, 2021, doi: 10.22059/jitm.2021.80026.
[32] A. Cuzzocrea, "Evolving OLAP and BI towards complex, high-performance big-OLAP-data-cube-processing analytics frameworks: How to speed-up large-scale, high-dimensional queries over clouds," *Procedia Computer Science*, vol. 246, pp. 4169–4175, 2024, doi: 10.1016/j.procs.2024.09.256.
[33] A. M. Deshmukh and R. Chalmeta, "Validation of system usability scale as a usability metric to evaluate voice user interfaces," *PeerJ Computer Science*, vol. 10, Feb. 2024, doi: 10.7717/peerj-cs.1918.
[34] N. Qatawneh, "Empirical insights into business intelligence adoption and decision-making performance during the digital transformation era: Extending the TOE model in the Jordanian banking sector," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 4, Dec. 2024, doi: 10.1016/j.joitmc.2024.100401.
[35] G. Dean, J. Moraes, J. White, R. Deleon, M. Jones, and T. Furlani, "Performance Optimization of the Open XDMoD Datawarehouse," in *Practice and Experience in Advanced Research Computing*, New York, NY, USA: ACM, Jul. 2022, pp. 1–7, doi: 10.1145/3491418.3530290.
[36] R. Salgotra, P. Sharma, S. Raju, and A. H. Gandomi, "Correction: A Contemporary systematic review on meta-heuristic optimization algorithms with their MATLAB and Python code reference," *Archives of Computational Methods in Engineering*, Dec. 2024, doi: 10.1007/s11831-024-10206-3.
[37] L. M. Hoi, W. Ke, and S. K. Im, "Manipulating data lakes intelligently with java annotations," *IEEE Access*, vol. 12, pp. 34903–34917, 2024, doi: 10.1109/ACCESS.2024.3372618.
[38] R. Fitriana, D. Sugiarto, M. K. Hidayat, and Y. E. H. Turnip, *Design of a business intelligence system model*. Wawasan Ilmu, 2024.
[39] A. Popović, V. Ivković, N. Trajković, and I. Luković, "A domain-specific language for managing ETL processes," *PeerJ Computer Science*, vol. 10, Jan. 2024, doi: 10.7717/peerj-cs.1835.
[40] P. Yin and J. Cheng, "A MySQL-based software system of urban land planning database of Shanghai in China," *Computer Modeling in Engineering & Sciences*, vol. 135, no. 3, pp. 2387–2405, 2023, doi: 10.32604/cmes.2023.023666.

BIOGRAPHIES OF AUTHORS

Miwan Kurniawan Hidayat has an educational background as a graduate of Informatics Engineering at Pasundan University, Bandung, a master of Computer Science graduate at Nusa Mandiri University, Jakarta, and a graduate of a Master of Industrial Engineering at Trisakti University, Jakarta. Career as an IT practitioner and lecturer at Bina Sarana Informatika University, Jakarta in the field of Information Engineering. Many papers have been published in scientific journals. He can be contacted at email: miwan@bsi.ac.id.

Dedy Sugiarto is an associate professor and lecturer in the Department of Information Systems at Universitas Trisakti, Indonesia. He pursued his Doctoral Degree in Agro-Industrial Engineering at IPB University. His research interests include knowledge management, applied statistics, and data mining and text mining. He can be contacted at email: dedy@trisakti.ac.id.



Rina Fitriana is an associate professor and lecturer in the Department of Industrial Engineering, Universitas Trisakti, Indonesia. Her educational background is a Bachelor of Industrial Engineering from Universitas Trisakti, a Magister of Management from PPM School of Management, and a Doctor of Agro-Industrial Technology from IPB University. She has more than 22 years of teaching/research in the field of Industrial Engineering. Her research interests include quality engineering, analyze system analysis of information system design, data mining, and business intelligence. She can be contacted at email: rinaf@trisakti.ac.id.

Yun-Chia Liang is professor and chair, Department of Industrial Engineering and Management, Yuan-Ze University, Taiwan, ROC. His research and interests are meta-heuristics, intelligent computing, production scheduling, logistics management, artificial neural network applications. He can be contacted at email: ycliang@saturn.yzu.edu.tw.