Journal of Emerging Information Systems and Business Intelligence
ISSN: 2774-3993
Journal homepage: https://ejournal.unesa.ac.id/index.php/JEISBI/

Healthcare Data Analysis Through Business Intelligence: A Case Study with Power BI

Avikatria Cahyaningrum[1], I Kadek Dwi Nuryana[2]

[1, 2] *Surabaya State University, Surabaya, Indonesia*

avikatria.22151@mhs.unesa.ac.id, dwinuryana@unesa.ac.id

ABSTRACT

This research aims to analyze hospital health data with the application of Power BI-based Business Intelligence (BI) to support a more precise and efficient decision-making process. The research data is taken from a public repository that provides a hospital management system database structure with complex inter-table relationships. The initial stages were carried out with the ETL (Extract, Transform, Load) process to integrate and clean the data before being entered into the data warehouse with the star and galaxy schema model. Next, analysis was conducted using Online Analytical Processing (OLAP) for medical service usage and other trends. In addition, the application of data mining using the Random Forest algorithm is also carried out for the classification of hospital busyness levels and prediction of patient re-visits based on historical data.

Keywords: Business Intelligence, ETL, OLAP, Data Mining, Random Forest, Power BI, Dashboard

Article Info:
Article history:
Received July 25, 2025
Revised August 16, 2025
Accepted November 15, 2025

Corresponding Author
Avikatria Cahyaningrum
Surabaya State University, Surabaya, Indonesia
avikatria.22151@mhs.unesa.ac.id

1. INTRODUCTION

Hospitals face challenges in managing complex and unstructured big data, especially in supporting fast and accurate decision making. Inefficiencies in processing data can have an impact on service quality and delays in management response.

Business Intelligence (BI) comes as a solution to manage and analyze health data systematically. With the help of ETL (Extract, Transform, Load) process, data from hospital management systems can be processed and stored in a data warehouse, then analyzed using OLAP (Online Analytical Processing) techniques and classification algorithms. This approach enables accurate data-driven decision-making.

Previous research has proven the effectiveness of BI in the context of healthcare. For example, Iswara et al [1] applied OLAP to monitor quality indicators of hospital services, while Asaury et al. [3] used the Random Forest algorithm to predict the number of admissions. However, most of these studies have not integrated all stages of BI thoroughly from ETL, data warehouse, OLAP, to data mining and dashboard visualization in one integrated system.

This research develops a Power BI-based BI system that integrates the ETL stage, data storage in a data warehouse using star and galaxy schemas, OLAP analysis, and the application of data mining using the Random Forest algorithm. The results of this analysis are visualized in the form of an interactive dashboard to support decision making by hospital management.

The purpose of this research is to build an interactive dashboard system that is able to provide important information visually and assist hospital management in making quick and precise decisions. This research is expected to contribute to the application of BI in the health sector, especially in improving service efficiency and data-based hospital management strategies.

2. METHODS

This research is an applied research with a quantitative approach that aims to develop a Business Intelligence (BI) system in analyzing hospital service data based on a data warehouse. This architecture has 3 stages, namely Data Sources, Data Flow and BI Results. The flow of the architecture is shown in Figure 1.

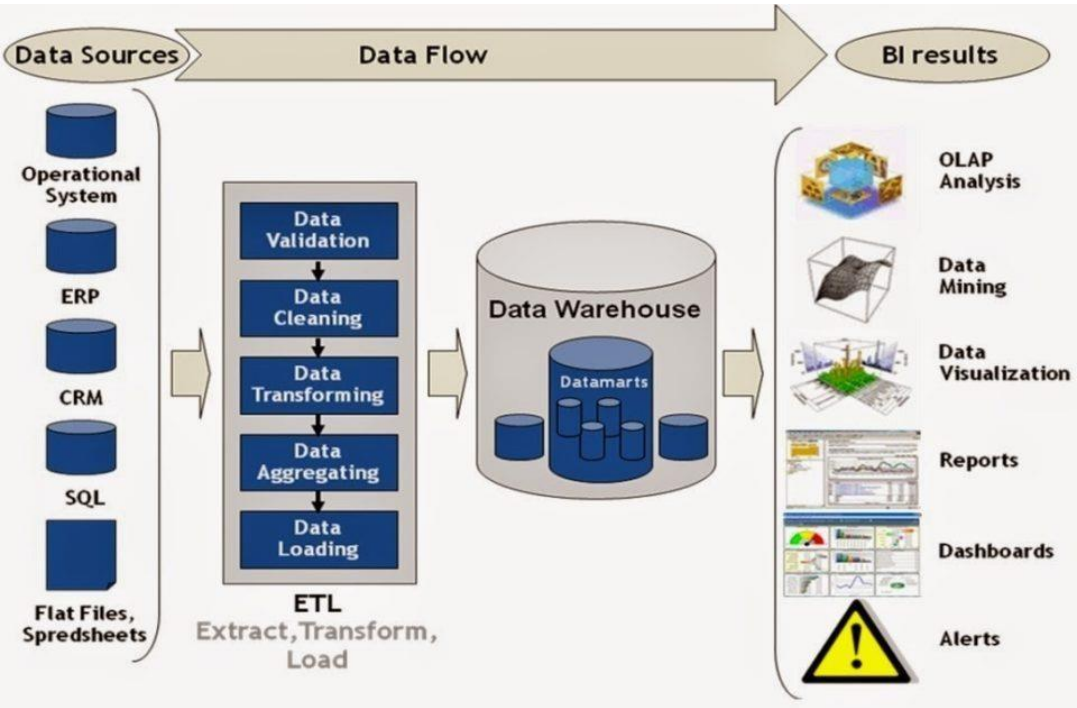

Figure 1. Business Intelligence Architecture

2.1. Data Sources

Enterprise Resource Planning (ERP) is a concept for managing company resources in the form of an integrated software system designed to serve and support various functions within the company [3]. Customer Relationship Management (CRM) is a business plan to understand, anticipate, and manage the needs of potential customers in an organization at this time and in the future [1]. Databases use SQL commands to modify data in tables, so SQL can easily describe the relationship between tables and rows [7].

2.2. Data Flow

The Extract process is data extracted from available internal and external sources [5]. In the extraction process there are two stages data validation is used to maintain the authenticity of data sent by users into the application [6] and data cleaning, at this stage when the information obtained is not really needed, has repeated data, inconsistent data and incorrect data, it will be reviewed and cleaned [4].

Transform is the process of converting data that was originally in operational form so that it can be used as a data warehouse format [6]. In the transform process there are data transforming is used to adjust data into a new scale so that it is eligible for analysis [8] and data aggregation is a relationship that describes all parts of the relationship and is commonly referred to as a relationship [9]. Load is the stage used to enter data into a data warehouse [7].

The data warehouse is used as a data storage area from various sources for certain purposes in order to speed up data access [10]. Creating a data warehouse can use the Kimball Model, this model has nine steps in creating a data warehouse [11]. Choose the process, selecting business processes that are suitable and can provide solutions to problems that occur. Choose the grain, grain or data granularity level is data from candidate facts that can be analyzed. Grain describes how detailed data is recorded, and is an important foundation in dimensional schema design. Identify and conform the dimensions, at this stage, identification and customization with the fact table will be carried out. The dimension table will contain a collection of dimensions or viewpoints that describe the facts in the fact table. Choose the facts, this stage is used to select the facts that will be used. Each fact has data that can be calculated to be displayed in the form of reports, graphs or diagrams. Store precalculations in the fact table, the selected facts will be reviewed with the aim of determining the opportunity to pre-calculate. Calculations on an attribute should be considered to be stored in the fact table, in order to reduce the risk of errors in the program every time it performs calculations on that attribute. Round out the dimension tables, facts that have been determined based on the results of identifying dimensions are then made descriptions that contain structured information about the attributes in the dimension table. Dimension tables must be fully annotated and easy to understand. Choose the duration of the database, at this stage is the determination of the duration of time that will be stored in the data warehouse. Determine to track slowly changing dimensions, this stage is used to track slow dimensional changes, by rewriting the changed attribute, creating new records on the dimension, and creating an alternative attribute that holds the new value. Decide the physical design, at this stage, the implementation of the physical design is carried out by following the eight steps above. So that a good data warehouse will be formed.

2.3. BI Result

Business Intelligence or commonly abbreviated as BI is a set of analytical technologies that can help store, process, analyze, and access big data. Business Intelligence systems can also recognize market trends and can describe graphic images based on the data accessed [12].

Online Analytical Processing is used to filter the information stored in the database in an easy way [10]. OLAP has a several operation such as, roll up, an operation used to extract data from a lower level to a higher level. Drill down, an operation used to

extract data from a higher level to a lower level. Slice and dice, an operation used to analyze data from different points of view by taking "chunks" of data and analyzing them separately. Pivot, an operation used to change the orientation of data from one dimension to another. Dicing, an operation similar to slice and dice, but used to analyze data from multiple dimensions at once. Drill through, an operation used to view the underlying details of the data being analyzed. Rollup and drill through, operations used to extract data from lower levels to higher levels and vice versa. OLAP has the ability to analyze data multidimensionally, there are three things that need to be considered, dimension an important component of OLAP that is used to determine the right point of view in analyzing data. Measures, units used to measure data. Time, units of time used in data analysis.

Data mining is the process of finding information in large databases [13]. Classification is the assignment of objects to one of several predefined categories. The input data for classification is a collection of records. Each record is known as an instance, which is defined by a tuple (x, y) where x is the set of attributes and y is a particular attribute, expressed as a class label or as a target attribute [14]. The random forest method is a method that can improve accuracy results because in generating anal nodes for each node is done randomly. This method is used to build a decision tree consisting of root nodes, internal nodes and leaf nodes by taking attributes and data randomly according to the conditions imposed. The root node is the topmost node, or commonly referred to as the root of the decision tree. Internal node is a branching node, where this node has a minimum of two outputs and only one input. While the leaf node or terminal node is the last node that has only one input and no output.

The visualization used can make it easier for users to describe the presentation of data so that it looks attractive and interactive [10]. Dashboard is a part of Power BI that functions to display data along with its visualization [9].

3. RESULTS AND DISCUSSION

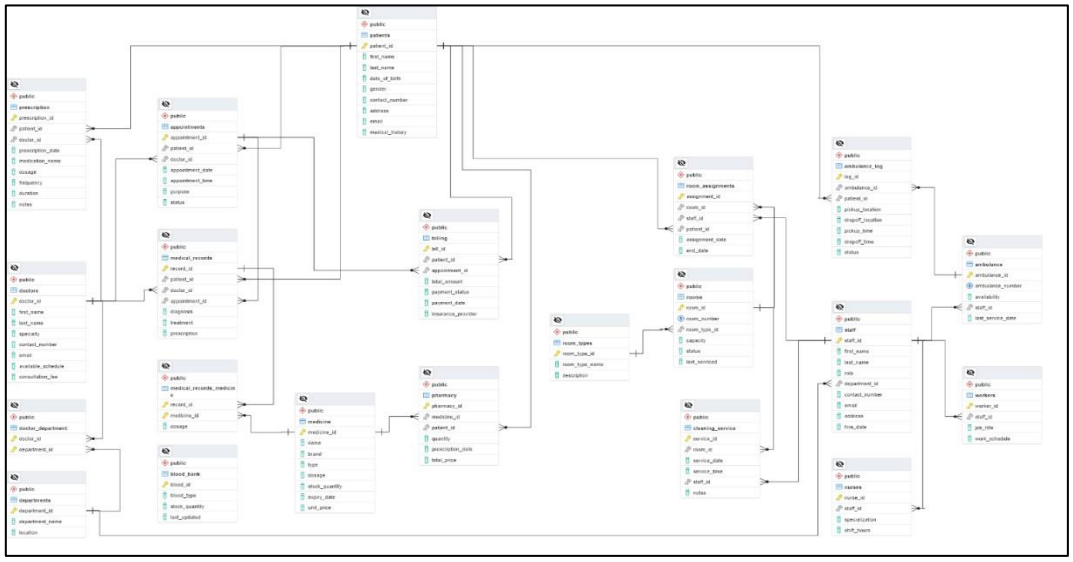

Figure 2. Entity Relationship Diagram

The data used in this study is stored in a PostgreSQL database consisting of 21 interconnected relational tables, as shown in Figure 3.6. This database structure is adopted

from Github which provides a relational design. Each table contains manually and automatically generated dummy data to realistically simulate hospital operations. The following is an explanation of the table:

Table 1. stores personal information about the patient and stores the patient's medical history. This table has 9500 data.

Table 1. Patients Table

No	Column Name	Data Type	Character Length
1)	patient_id	Integer	-
2)	first_name	Varying Character	50
3)	last_name	Varying Character	50
4)	date_of_birth	Date	-
5)	gender	Character	-
6)	contact_number	Varying Character	15
7)	address	Varying Character	255
8)	email	Varying Character	100
9)	medical_history	Text	-

Table 2. contains information about doctors such as, doctor specialization, practice schedule and contact information. This table has 30 data.

Table 2. Doctors Table

No	Column Name	Data Type	Character Length
1)	doctor_id	Integer	-
2)	first_name	Varying Character	50
3)	last_name	Varying Character	50
4)	specialty	Character Varying	100
5)	contact_number	Varying Character	15
6)	email	Character Varying	100
7)	available_schedule	text	-
8)	consultation_fee	integer	-

Table 3 stores a list of departments in the hospital such as department names and department locations. This table has 11 data.

Table 3. Table departments

No	Column Name	Data Type	Character Length
1)	department_id	Integer	-
2)	department_name	Varying Character	50
3)	location	Character Varying	100

Table 4. serves to connect the doctor table and the department table. This table has 30 data.

Table 4. Doctor_departments table

No	Column Name	Data Type
1)	doctor_id	Integer
2)	department_id	Integer

Table 5. is used to schedule patient appointments that store plans, status and doctor referrals. This table has 39116 data

Table 5. Appointments table

No	Column Name	Data Type	Character Length
1)	appointment_id	Integer	-
2)	patient_id	Integer	-
3)	doctor_id	Integer	-
4)	appointment_date	Date	-
5)	appointment_time	Time	-
6)	purpose	Varying Character	255
7)	status	Varying Character	20

Table 6. is used to store patient medical records that store patient diagnoses, treatments and drug prescriptions that are related to the appointments table. This table has 39116 data.

Table 6. Medical_records table

No	Column Name	Data Type
1)	record_id	Integer
2)	patient_id	Integer
3)	doctor_id	Integer
4)	appointment_id	Integer
5)	diagnosis	Text
6)	treatment	Text
7)	prescription	Text

Table 7 is used to store information about drugs that can display drug names, doses, expirations, and drug types. This table has 44 data.

Table 7. Medicine table

No	Column Name	Data Type	Character Length
1)	medicine_id	Integer	-
2)	name	Varying Character	100
3)	brand	Character Varying	50
4)	type	Character Varying	20
5)	dosage	Character Varying	50
6)	stock_quantity	Integer	-
7)	expiry_date	Date	-
8)	unit_price	Integer	-

Table 8. combines the medical records and medicine tables. This table has 39116 data.

Table 8. Medical_records_medicine table

No	Column Name	Data Type	Character Length
1)	record_id	Integer	-
2)	medicine_id	Integer	-
3)	dosage	Varying Character	50

Table 9. is used to store prescriptions that doctors give to patients and details the treatment such as dosage, frequency and duration of drug use. This table has 39116 data.

Table 9. Prescription table

No.	Column Name	Data Type	Character Length
1)	prescription_id	Integer	-
2)	patient_id	Integer	-
3)	doctor_id	Integer	-
4)	prescription_date	Date	-
5)	medication_name	Varying Character	100
6)	dosage	Character Varying	100
7)	frequency	Character Varying	50
8)	duration	Character Varying	50
9)	notes	Character Varying	255

Table 10 is used to record drugs prescribed to patients by connecting the drug and patient tables with quantity and date. This table has 39116 data.

Table 10. Pharmacy table

No.	Column Name	Data Type
1)	pharmacy_id	Integer
2)	medicine_id	Integer
3)	patient_id	Integer
4)	quantity	Integer
5)	prescription_date	Date
6)	total_price	Integer

Table 11 is used to serve payment information that displays the total cost, payment status and insurance provider. This table has 39116 data.

Table 11. billing table

No	Column Name	Data Type	Character Length
1)	bill_id	Integer	-
2)	patient_id	Integer	-
3)	appointment_id	Integer	-
4)	total_amount	Numeric	-
5)	payment_status	Varying Character	20
6)	payment_date	Date	-
7)	insurance_provider	Varying Character	100

Table 12. is used to define the type of room in the hospital. This table has 15 data.

Table 12. room_types table

No	Column Name	Data Type	Character Length
1)	room_type_id	Integer	-
2)	room_type_name	Varying Character	50
3)	description	Varying Character	255

Table 13. is used to track room types in detail. This table has 42 data.

Table 13. Rooms table

No	Column Name	Data Type	Character Length
1)	room_id	Integer	-
2)	room_number	Varying Character	10
3)	room_type_id	Integer	-
4)	capacity	Integer	-
5)	status	Varying Character	20
6)	last_serviced	Date	-

Table 14. is used to record the assignment period with reference to rooms, staff and patients. This table has 39116 data.

Table 14. rooms_assignment table

No	Column Name	Data Type	Character Length
1)	room_id	Integer	-
2)	room_number	Varying Character	10
3)	room_type_id	Integer	-
4)	capacity	Integer	-
5)	status	Varying Character	-
6)	last_serviced	Date	-

Table 15. is used to store hospital staff roles such as nurses, and staff with different departments. This table has 70 data.

Table 15. Staff table

No.	Column Name	Data Type	Character Length
1)	staff_id	Integer	-
2)	first_name	Varying Character	50
3)	last_name	Varying Character	50
4)	role	Character Varying	20
5)	department_id	Integer	-
6)	contact_number	Varying Character	15
7)	email	Character Varying	50
8)	address	Text	-
9)	hire_date	Date	-

Table 16. is used to store information on staff whose specialty is nursing and can display their working hours and departments. This table has 30 data.

Table 16. Nurses table

No.	Column Name	Data Type	Character Length
1)	Nurse_id	integer	-
2)	Staff_id	integer	-
3)	specialization	Character Varying	50
4)	Shift_hours	text	-

Table 17. used to store non-medical staff information. This table has 9 data.

Table 17. Workers Table

No.	Column Name	Data Type	Character Length
1)	worker_id	Integer	-
2)	staff_id	Integer	-
3)	job_tittle	Varying Character	50
4)	work_schedule	Text	-

Table 18. is used to track room cleaning activities. This table has 30702 data.

Table 18. cleaning_service table

No	Column Name	Data Type	Character Length
1)	service_id	Integer	-
2)	integer	Integer	-
3)	service_date	Date	-
4)	service_time	Time	-
5)	staff_id	Integer	-
6)	notes	Varying Character	255

Table 19. is used to check availability, service dates and driver assignments. This table has 5 data.

Table 19. Ambulance table

No.	Column Name	Data Type	Character Length
1)	ambulance_id	Integer	-
2)	ambulance_number	Varying Character	10
3)	availability	Character Varying	15
4)	staff_id	Integer	-
5)	staff_service_date	Date	-

Table 20. used to record the use of ambulances for patient transportation. This table has 500 data.

Table 20. Ambulance_log table

No.	Column Name	Data Type	Character Length
1)	log_id	Integer	-
2)	ambulance_id	Integer	-
3)	patient_id	Integer	-
4)	pickup_location	Varying Character	100
5)	dropoff_location	Character Varying	100
6)	pickup_time	Time	-
7)	dropoff_time	Time	-
8)	status	Character Varying	15

Table 21. used to manage blood type stocks. This table has 10 data.

Table 21. blood_bank table

No	Column Name	Data Type	Character Length
1)	blood_id	Integer	-
2)	blood_type	Varying Character	3
3)	stock_quantity	Integer	-
4)	last_updated	Date	-

CONCLUSION

This research shows that the application of Power BI-based *Business Intelligence* (BI) as a whole starting from the ETL process, designing *data warehouses* with star and galaxy schemes, OLAP analysis, to classification using the *Random Forest* algorithm can support the decision-making process in hospitals efficiently. The analysis results will be visualized through an interactive *dashboard* that provides important insights for hospital management.

This system can make a real contribution to data-driven hospital management and open up further development opportunities for wider scale implementation in the health sector.

REFERENCES

[1] A. W. Iswara, H. Setiadi, and A. Wijayanto, "Implementation of Business Intelligence for Quality Support of RSUD Ir. Soekarno Sukoharjo with Data Warehouse," 2020.

[2] A. S. Asaury, A. Hamid, and G. Triyono, "Prediksi Jumlah Pasien Masuk Rumah Sakit Menggunakan Metode Random Forest," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 2, Feb. 2025, doi: 10.52436/1.jpti.660.

[3] D. A. N. Wulandari, S. Anggraeni, and Imron, *Enterprise Resource Planning (ERP)*. Yogyakarta: Graha Ilmu, 2021.

[4] R. Oktapiani and T. Kusnadi, *Perancangan Sistem Basis Data dan Penerapannya dalam Sistem Informasi*. Yogyakarta: Teknosain, 2021.

[5] A. Prahendratno *et al.*, *Business Intelligent (Pengantar Business Intelligence dalam Bisnis)*. Jambi: Sonpedia.com, 2023.

[6] I. Amirulloh and H. Sutisna, *Mahir Framework Kekinian Tanpa Ribet; Studi Kasus Aplikasi Restaurant (Table Service)*. Yogyakarta: Graha Ilmu, 2020.

[7] I. D. B. G. K. J. Nugraha, A. Susanto, and A. S. Girsang, "Perancangan Data Warehouse dan Dashboard PT. Jaya Teknik," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 1, pp. 17–24, Apr. 2022, doi: 10.21063/jtif.2022.v10.1.17-24.

[8] A. A. J. Sinlae, *Keputusan Berbasis Data Panduan Praktis Data Science untuk Bisnis*. Malang: Literasi Nusantara, 2023.

[9] S. Masripah, N. Hidayati, N. O. Syamsiah, and Haryani, *Analisa Perancangan Sistem Informasi Akuntansi*. Yogyakarta: Graha Ilmu, 2019.

[10] I. Ahmad, Y. Rahmanto, and R. I. Borman, *Business Intelligence*. Bandarlampung: Universitas Teknorat Indonesia, 2023.

[11] G. S. Mahendra *et al.*, *Buku Ajar Sistem Pendukung Keputusan*. Jambi: Sonpedia, 2023.

[12] R. T. Yunardi and N. Z. Dina, *Data Mining dan Machine Learning dengan Orange3 Tutorial dan Aplikasinya*. Surabaya: Airlangga University Press, 2022.

[13] N. Iriadi, Priatno, and A. Ishaq, *Penerapan Data Mining dengan Rapid Miner; Konsep Data Maining, Data Warehouse, Metode, Model, Teknik*. Yogyakarta: Graha Ilmu, 2020.

[14] Rulin Swastika, Siti Mukodimah, Ferry Susanto, Muhamad Muslihudin, and Sri Ipnuwati, *Implementasi Data Mining (Clastering, Association, Prediction, Estimation, Classification)*. Indramayu: Penerbit Adab, 2023.
[15] I. Putikadea, A. Kusumaningsih, M. D. Yanthi, and C. S. Siregar, *Manajemen Basis Data*. Yogyakarta: Bintang Semesta Media, 2022.