International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 5, October 2024, pp. 5612~5625
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i5.pp5612-5625

Real-time business intelligence development using machine learning to increase the potential of the dairy goat milk business

Alusyanti Primawati[1,2], Imas Sukaesih Sitanggang[1], Annisa[1], Dewi Apri Astuti[3]
[1]Department of Computer Science, IPB University, Bogor, Indonesia
[2]Informatics Engineering Study Program, Faculty of Engineering and Computer Science (FTIK), Universitas Indraprasta PGRI, Jakarta, Indonesia
[3]Department of Nutrition Science and Feed Technology (INTP), Faculty of Animal Husbandry, IPB University, Bogor, Indonesia

Article Info

Article history:

Received Apr 22, 2024
Revised Jun 25, 2024
Accepted Jul 2, 2024

Keywords:

Dairy goat milk
Long short-term memory
Machine learning
Real-time business intelligence
RTBI frameworks

ABSTRACT

The development of big data and real-time data warehouse (RTDW) technologies has transformed traditional business intelligence (BI) into real-time business intelligence (RTBI). The RTBI framework is developed in this study by incorporating machine learning-based real-time prediction features. The complexity of layer integration in the RTBI framework is a challenge in building RTBI. The development of RTBI was carried out in business areas that did not have RTBI from the beginning, such as the dairy goat milk business in Probolinggo, East Java. Another main reason is that the dairy goat milk business is a food alternative to cow's milk in Indonesia. The results of this study can contribute to increasing the potential value of the goat milk business. The research method was developed by adapting to the Kimball method and unified modeling language (UML). The real-time prediction feature with the long short-term memory (LSTM) algorithm is the main feature in the RTBI framework developed in the research. The calculation results of real-time predictive analysis latency successfully approached 0 milliseconds (ms), namely 9.35×10^{-5} ms. The application of RTBI in the dairy goat milk business was successfully built but the real data is very limited, so RTBI is less able to describe the movement of the business.

Corresponding Author:

Alusyanti Primawati
Department of Computer Science, IPB University
Bogor, Indonesia
Email: 1149alusyanti@apps.ipb.ac.id

1. INTRODUCTION

Business intelligence (BI) innovations and applications provide great support to managers to make the right decisions at the right time [1], [2]. The advantages of BI are recognized by many organizations as the set of visual tools resulting from BI analysis are used for business benefits [1]. BI is a tool to process data starting from data sources processed into a data warehouse (DW), data exploration, data mining with business analytics tools, and finally data visualization into a BI dashboard [3]–[5]. However, the right decision at the right time requires the right analysis model with data processing and analysis latency near 0 milliseconds (ms). Traditional BI has the disadvantage of high latency of processing data into DW and high latency of analysis because analysis cannot be done immediately due to waiting for data refresh from DW [4]. The disadvantages of DW are overcome through the presence of a real-time data warehouse (RTDW) by developing extract, transform, and load (ETL) techniques that are close to real-time [6]–[8]. The development of DW into RTDW [9], [10] as well as the existence of streaming data streams from big data technology has enabled BI to evolve into real-time business intelligence (RTBI).

The analytics layer in BI is already evolving rapidly due to the emergence of machine learning integration trends and opportunities in BI [11]. The integration of machine learning in BI has given rise to predictive analytics as the second of three levels of business analytics [1], [12]. Machine learning (ML) is known to excel at finding patterns and trends in large historical data, so BI integrated with ML can make accurate predictions about the future [12]. Real-time data is characterized by time series data, a popular ML model for time series prediction is long short-term memory (LSTM) [13]–[20]. Thus, ML models in BI tools can maximize operations and improve business results by analyzing variables and making data-driven decisions [21]. Data-driven decisions are expected to increase business potential. However, analytic modules built from historical datasets need an effective DW schema [10], [22]. The right DW scheme can also answer the challenges of big data because the data stored in DW has been processed to be structured. The emergence of big data due to rapidly growing internet services raises new challenges related to data volume, data heterogeneity, data variation, processing latency, and data accuracy [23]. The determination of the right DW model scheme is not easy because the DW design process is also related to the data model that will be used in answering business questions needed by business people. For example, supply chain management requires answers to questions about how the current production and distribution trends (descriptive analysis) and whether future production and distribution will increase (predictive analysis) and others. Quick answers to business questions at the right time from the results of business analysis can accelerate determining strategies related to business profitability and investor revenue and profits [1].

The previous RTBI framework consisted of 4 main layers: operational business layer, real-time data processing layer, real-time analytics layer, and presentation layer [7]. Today's operational business layer has evolved into real-time as the phenomenon of big data technology has created many real-time operational business tools such as online applications [24]. However, the RTBI architecture designed is only for data analysis other than RTDW using machine learning-based classification models. If the business needs require predictive analysis results from RTDW, the architecture needs to be modified according to business needs. In addition, the analysis is not done in real time. The research gap found that the RTBI framework needs to be modified for real-time prediction needs using ML accompanied by an integrated process between layers as a real-time mechanism. The resulting RTBI system architecture and mechanism are different from previous researchers due to different business needs and the use of different technologies at each layer and need one additional layer to integrate in real-time each layer. Therefore, a reality environment that does not have BI from the beginning is needed to develop an RTBI framework to measure analysis latency when machine learning models are implemented into RTBI applications.

The RTBI framework is modified by adding a real-time integration system layer with a bi-directional connection between the real-time data processing layer and the real-time visualization layer as shown in Figure 1. The integration sub-layer also houses the real-time data retrieval mechanism from the RTDW server to be analyzed in the real-time analytics layer. The complexity of integrating the four layers of RTBI is a challenge in building RTBI from scratch [8]. The right RTBI design and implementation can provide fast business information at the right time. Therefore, it is important to create a practical integrated process to reduce the complexity of designing machine learning-based BI applications with real-time technology. The design process starts with designing an effective conceptual data model to incorporate the model features used by each RTBI layer [25]. An effective conceptual model is fundamental to the success of a BI project [25]. The conceptual model was built based on the RTBI framework developed from previous studies [7].

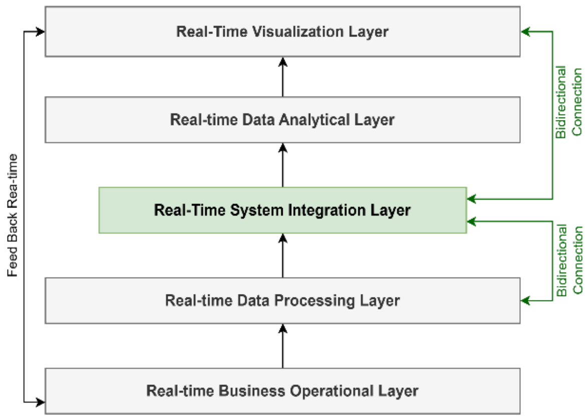


Figure 1. The RTBI framework developed in this study

This research focuses on how to design an RTBI application for a business that has not had BI since its inception based on the RTBI framework proposed in Figure 1. The reality environment in this research road map is the dairy goat milk business in East Java. Dairy goat farming in Indonesia needs RTBI to increase business value and become a leading commodity [26]. The RTBI application can support precision livestock farming (PLF) with an overview of the potential business value of goat milk because fresh milk products from dairy goat farms are one of the food alternatives after cow's milk, therefore a descriptive and predictive analysis model is needed that can present trends and predictions of future goat fresh milk production using machine learning [27], [28]. The BI conceptual model was built first to build the RTBI of dairy goat farming with a machine-learning approach.

A simple and data-centered conceptual model will be designed in the RTBI project with a case study of a dairy goat farm [4], [25]. The advantage is that the researcher can focus on the type of variables and the process of filtering or aggregating the data used according to the objectives of the RTBI project. The initial preparation starts with building the traditional dairy goat farm BI and the models used in DW, descriptive analytics, and predictive and dashboard BI templates developed into RTBI. Traditional BI modeling using Kimball method, RTBI with unified modeling language (UML) [4], [29]. The scope of this research is the data used in the business process of dairy goat milk production and distribution in Probolinggo, East Java. Furthermore, chapter 2 of this paper discusses the methods used in designing BI and developed into RTBI, the results and discussion are presented in chapter 3, and chapter 4 contains conclusions from the discussion results along with recommendations for future research.

2. METHOD

2.1. Experimental materials

Real-time business intelligence was built using data from the dairy goat milk business in Probolinggo, East Java. The materials required in the research are OLTP database sources from real-time data collection tools and the PostgreSQL database. The data collection tool uses an Android application-based smartphone. The application records various transactional data on livestock population, goat milk production, and distribution. Some of the tables used to build goat milk business intelligence are unit_livestock, region, milk production, milk distribution, product type, partner category, and business partner as shown in Figure 2. The data tables in Figure 2 are used to build the logical model of the data warehouse. The construction of logical and physical data models using ETL tools namely Pentaho version 9.3 and the results are stored in PostgreSQL 13. The source of milk production data was taken from the Puska database server and dairy goat farmers in Probolinggo in June 2023. The dataset from Puska was still small, so daily milk production data from Prof Farm (365 days) and Nyx Farm (166 days) were added to the local Puska database.

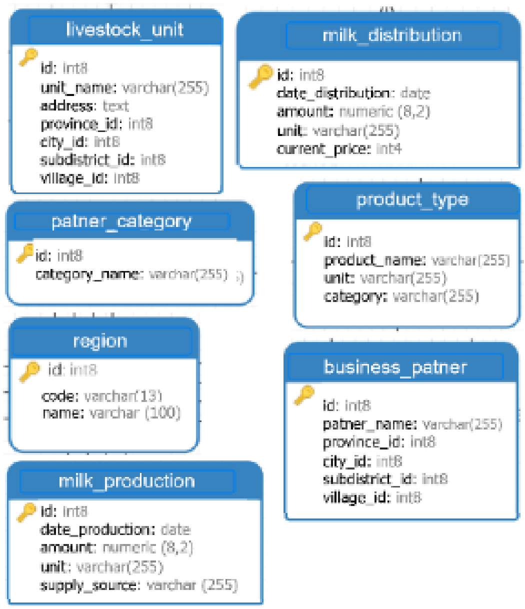


Figure 2. The data table is taken from the real-time operational business layer

The materials used to design multidimensional data models, traditional BI dashboards, and RTBI web applications include Power BI, Python with Jupyter for machine learning models, Visual Studio Code, Docker, FastAPI for real-time BI services (back-end), Vercel for RTBI dashboards (front-end), and 12th Gen Intel Core i5 specification laptops, 16 GB RAM, Windows 11 as compatible hardware.

2.2. Research methodology

This research methodology uses the Kimball methodology which focuses on designing conceptual models, logical models, physical models of multidimensional data, online analytical processing (OLAP) operations, and BI dashboard designs to be developed into RTBI dashboard prototypes [24]. RTBI design is presented with UML [4], [29]. The RTBI end-user application is integrated with a machine learning-based predictive model to predict the daily milk production of dairy goats in Probolinggo, East Java. The research stage consists of 6 main stages, namely preliminary stages, RTBI architecture design, traditional BI conceptual model for selecting the best multidimensional model scheme, traditional BI application design with Power BI for testing multidimensional model schemes, and website-based RTBI application design stages, RTBI system testing, deployment, and the final stage, namely conclusions and recommendations for RTBI development in similar case studies as shown in Figure 3.

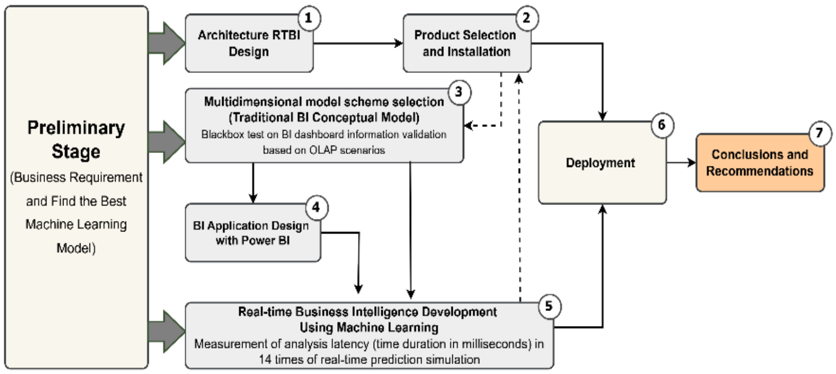


Figure 3. Research stage

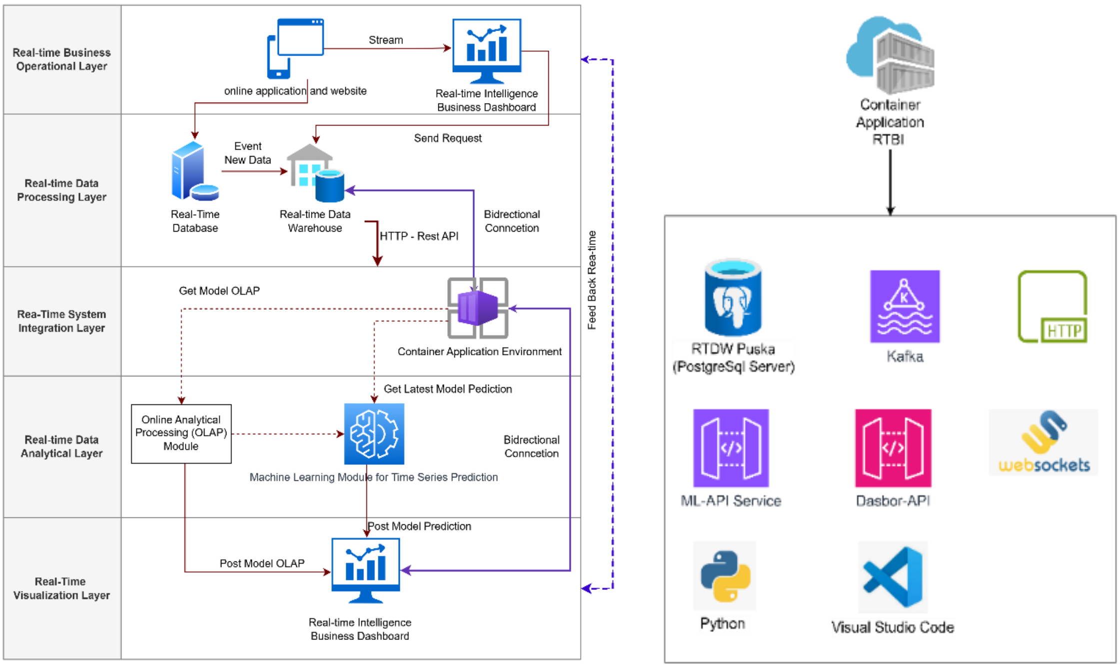
2.2.1. Preliminary stages

The first stage is the preliminary stage between the literature review [26] and the focus discussion group in 2022 to obtain information on the needs of the dairy goat milk business in Probolinggo, East Java. milk business needs information includes periodic milk production information, the amount of milk distributed, information on distributor partners who often request milk, feed information and the amount of milk produced, breeding information, and milk price information for each region. Trend information and prediction of goat milk production is important information that needs to be displayed to end-users because the potential picture of the goat milk business is seen through this information. The goat milk production prediction model is modeled using the best model studied previously, namely long short-term memory (LSTM) [19], [17]. LSTM has advantages in modeling time series history data [13]–[15], [18], [20].

2.2.2. RTBI architectural design, product selection and installation

The RTBI architecture was designed based on the RTBI framework developed for the case study in this research as shown in Figure 4(a). Each layer in the RTBI framework is determined by the technology product used for RTBI implementation. The real-time operational business layer is filled with smartphone or website-based online applications. The website-based RTBI dashboard is also part of the operational business layer as a user interface for user requests for business information. The real-time data processing layer uses RTDW as a server. The streaming data flow is built with Kafka. Data from RTDW is retrieved and sent through Kafka to be analyzed in the real-time analytics layer using mechanisms in the real-time system integration layer. The service mechanism with HTTP protocol and bidirectional connection in the real-time data integration layer is executed when new data enters RTDW. When the protocol HTTP is run, the OLAP module and ML module are run in the real-time analytics layer to analyze the latest historical data. All services are stored in containers. Docker is a container application environment used in the real-time system integration layer. Furthermore, the response results will be posted by Kafka to the real-time visualization layer by running the bi-directional protocol. The bidirectional connection protocol used in the RTBI architecture is WebSocket.

The products selected and installed in the RTBI architecture include Puska Apps as the real-time operational business layer, and Docker as the container application environment used in the real-time system integration layer. Puska RTDW for the real-time data processing layer, real-time OLAP module and real-time LSTM prediction module for the real-time analytics layer, and RTBI website as real-time visualization layer. All services are stored in Docker for easy deployment as shown Figure 4(b).



(a) (b)

Figure 4. The designed RTBI architecture (a) Technologies and modules used in each RTBI layer and (b) RTBI services and applications integrated into the container application environment

2.2.3. Multidimensional model schema selection (traditional BI conceptual model)

The third stage is the design of the BI conceptual model. The conceptual model includes the design of logical models and physical data models to build a dairy goat milk business intelligence business. The results of the next model are applied to real-time business intelligence of the dairy goat milk business. RTBI of the dairy goat milk business is integrated with the machine learning model designed earlier. The logical model design for dairy business DW is three dimensions and two fact tables. The three dimensions include time dimension, farm unit dimension, product dimension, and business partner dimension. The fact tables built are production facts and milk distribution facts. The logical DW model will be built with 2 scenarios of the galaxy model scheme. The star scheme is not used because there are 2 fact tables. Both logical model scenarios are described in Figures 5 and 6.

The logical model is transformed into a physical model using Pentaho version 9.3 through the extract, transform, and load (ETL) process. The ETL process builds 5 dimensions: supply source dimension, time dimension, location dimension, livestock unit dimension, product type dimension, and business partner dimension. The five dimensions are used to build 2 facts, namely *fact_production* as a distribution fact as shown in Figure 7 and *fact_distribution* fact as a distribution fact as shown Figure 8. The ETL results are stored in PostgreSQL 13 as a data warehouse server as shown in Figure 9.

The next step is to select the best data warehouse schema to generate information that can answer business questions. Each multidimensional data model schema was designed into a physical multidimensional data model using Power BI. Both multidimensional data model scenarios were tested using the same business intelligence dashboard. The test used the black-box method with 5 OLAP scenarios to validate the information displayed by the business question design as shown in Table 1. The model schema with all correct information validation was selected as the model schema used in the dairy goat milk business intelligence dashboard design.

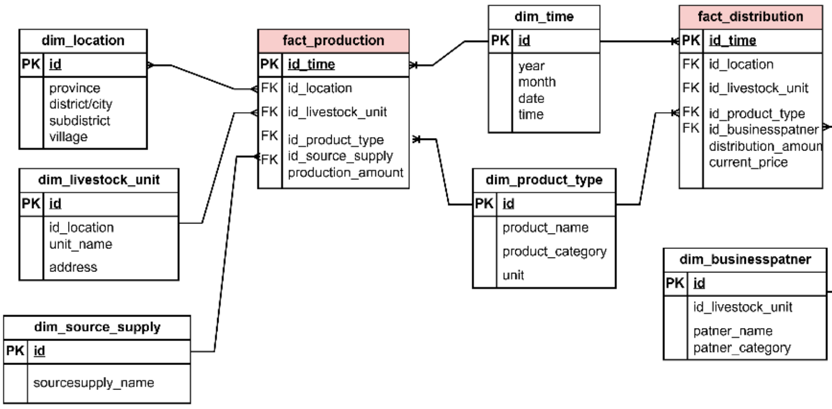


Figure 5. Schema model of galaxy scenario 1

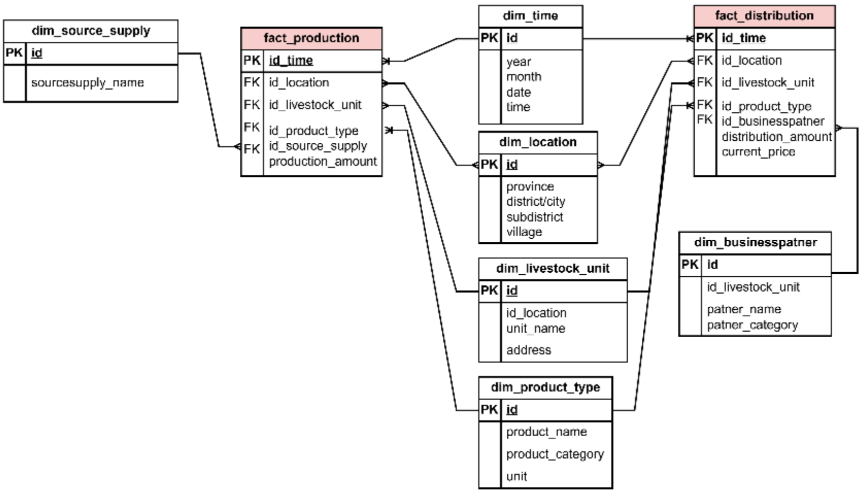


Figure 6. Schema model of galaxy scenario 2

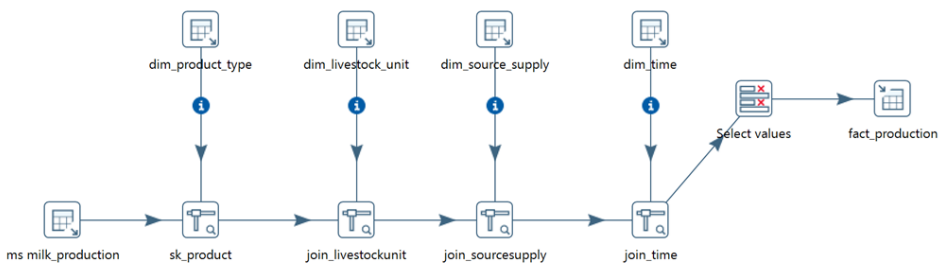


Figure 7. Production fact ETL process

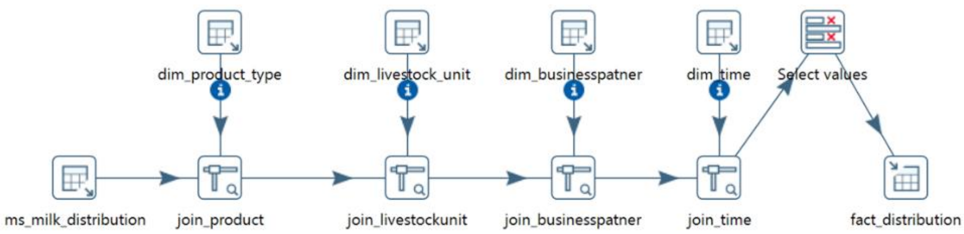


Figure 8. Distribution facts ETL process

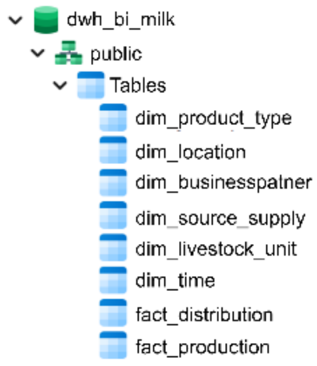


Figure 9. Dairy goat milk business data warehouse generated from the ETL process

Table 1. OLAP scenarios used for testing multidimensional model schemes

Scenario	Operation Description	Dimension Description
OLAP-1	Roll up to all	Time: all, Province: all, livestock unit: all
OLAP-2	Drill down and slice to year	Time: 2023, Province: all, livestock units: all
OLAP-3	Drill down and dice to year and province	Time: 2023, Province: East Java, livestock unit: all
OLAP-4	Drill down and dice to year, Province+Regency	Time: 2023, Province: East Java+Probolinggo district, livestock unit: all
OLAP-5	Drill down and dice to year, Province+Regency, and livestock units	Time: 2023, Province: East Java+Probolinggo district, Animal husbandry unit: Prof Farm

The BI dashboard design was required for the galaxy 1 and 2 schema trials. The BI dashboard was built using the Power BI tool. In the first row, the information presented is a summary of the total production and distribution of dairy goat milk products. In the second row, a plot of dairy goat fresh milk production and distribution is presented along with comparative information on total milk demand from business partners, distribution percentage, total revenue, average price, lowest price, and highest price of dairy goat fresh milk. Trend plots and predictions of goat milk production are presented based on the required information. Each visualized business information uses DAX Query as shown Table 2.

Table 2. Dax query design for traditional BI dashboard with Power BI

Query DAX *fact_produksi* Table	Function
dairymilk_production=CALCULATE(SUM(fact_production[production_amount]), fact_production[id_product_type]=3)	Calculate the amount of dairy milk production
production_pasteurization=CALCULATE(SUM(fact_production[production_amount]), fact_production[id_product_type]=4)	Calculate the amount of pasteurized milk production
production_kefir=CALCULATE(SUM(fact_production[production_amount]), fact_production[id_product_type]=5)	Calculate the amount of kefir production
production_yogurt=CALCULATE(SUM(fact_production[production_amount]), fact_production[id_product_type]=6)	Calculate the amount of yogurt production
Production_cheese=CALCULATE(SUM(fact_production[production_amount]), fact_production[id_product_type]=7)	Calculate the amount of cheese production
dis_dairymilk=CALCULATE(SUM(fact_distribution[distribution_amount]), fact_distribution[id_product_type]=3)	Calculate the amount of dairy milk distribution
dis_ pasteurization=CALCULATE(SUM(fact_distribution[distribution_amount]), fact_distribution[id_product_type]=4)	Calculate the distribution amount of pasteurized milk
dis_kefir=CALCULATE(SUM(fact_distribution[distribution_amount]), fact_distribution[id_product_type]=5)	Calculate the amount of kefir distribution
dis_yogurt=CALCULATE(SUM(fact_distribution[distribution_amount]), fact_distribution[id_product_type]=6)	Calculate the amount of yogurt distribution
dis_cheese=CALCULATE(SUM(fact_distribution[distribution_amount]), fact_distribution[id_product_type]=7)	Calculate the amount of cheese distribution
price_highest=Max(fact_distribution[current_price])	The highest fresh milk selling price
price_low=Min(fact_distribution[current_price])	The lowest selling price of fresh milk
price_rate=AVERAGE(fact_distribution[current_price])	Average Current selling price of dairy milk
*total_revenue=fact_distribution[amount_distribution]*fact_distribution[current_price]*	Total income
percentage_delivery=fact_distribution[dis_dairymilk]/fact_production[production_dairymilk]	Percentage Distribution

2.2.3. RTBI application design with machine learning approach, measurement of analysis latency, and deployment

Real-time system behavior on the RTBI dashboard layer is described using UML. The results of the dairy goat milk business BI dashboard design are applied to the design of the dairy goat milk business RTBI dashboard design. The prediction model in the goat milk business BI uses the forecasting feature of Power BI while the goat milk business RTBI is integrated with the predictive machine learning module, namely the LSTM model for milk prediction in Probolinggo, Prof Farm, and Nyx Farm. The real-time prediction workflow of LSTM is discussed separately in another paper. Measurement of analysis latency (time duration in milliseconds) in 14 times of real-time prediction simulation. Deployment is carried out by installing the RTBI application service docker image on the server as shown in Figure 4(b).

3. RESULTS AND DISCUSSION

3.1. Preliminary stages results

The results of the preliminary research conducted were a list of questions for the dairy goat milk business in East Java as shown in Table 3. Based on the results of focus group discussions with 2 units of dairy goat farmers in Probolinggo, East Java, researchers found 2 main topics needed by farmers, namely production and distribution topics. Thus, BI was built based on the main needs of production and distribution.

Table 3. Business information requirements in a case study of a dairy goat milk business in Probolinggo, East Java

No	Information	Business Question
1	Periodic milk production information	What is the current production of dairy milk, pasteurized milk, kefir, yogurt, and cheese in the East Java region? What is the current production of dairy milk, pasteurized milk, kefir, yogurt, and cheese in each district/city of East Java? What is the current production of dairy milk, pasteurized milk, kefir, yogurt, and cheese in each farm unit in each district/city of East Java? What is the trend of dairy milk production in the last year in East Java? What is the trend of dairy milk production in the past year in each district/city in East Java? What is the trend of dairy milk production in the past year in each farm unit in each district/city in East Java? What is the current percentage distribution of each type of dairy product in the East Java region? What is the current percentage distribution of each type of dairy product in each district/city of East Java? What is the current percentage distribution of each type of dairy in each farm unit in each district/city of East Java?
2	Information on the amount of milk distributed	What is the current distribution of dairy milk, pasteurized milk, kefir, yogurt, and cheese in the East Java region? What is the current distribution of dairy milk, pasteurized milk, kefir, yogurt, and cheese in each district/city of East Java? What is the current distribution of dairy milk, pasteurized milk, kefir, yogurt, and cheese in each farm unit in each district/city of East Java? What is the current distribution of dairy milk, pasteurized milk, kefir, yogurt, and cheese in the East Java region? What is the current distribution of dairy milk, pasteurized milk, kefir, yogurt, and cheese in each district/city of East Java? What is the current distribution of dairy milk, pasteurized milk, kefir, yogurt, and cheese in each farm unit in each district/city of East Java?
3	Periodic milk production information and amount of milk distributed	How is the trend comparison of dairy milk production and distribution in the last 1 year in the East Java region? How is the trend of production and distribution of dairy milk in the last 1 year in each district/city in East Java compared? How does the trend of dairy milk production and distribution in the last 1 year compare for each farm unit in each district/city in East Java? What is the percentage of milk distribution to current milk production in the East Java region? What is the percentage of milk distribution to current milk production in each district/city in East Java? What is the percentage of milk distribution to current milk production in each farm unit in each district/city in East Java?
4	Information on distributor partners who frequently request milk	How does the demand for milk from all current business partners compare across the East Java region? How does the demand for milk from all current business partners in each district/city in East Java compare? How does the current demand for milk from all business partners compare for each farm unit in each district/city in East Java? Name the region in East Java that has the most current demand for milk for distribution. Name the business partner that has the most demand for milk in each region in East Java at the moment.
5	Milk price information for each region	What is the average milk price in East Java? What is the average milk price in each district/city in East Java? What is the lowest milk price in East Java? What is the lowest milk price in each district/city in East Java? What is the highest milk price in East Java? What is the highest milk price in each district/city in East Java?

On the topic of production, farm units need information on predicting the amount of milk production in the future. Researchers have conducted initial experiments by selecting the best machine learning-based prediction model for milk production time series datasets and the LSTM model was selected as the best model as shown in Table 4 [18]. The LSTM model built is a daily prediction model using lockback = 2. There are 3 models built in the initial experiment, namely prediction models for milk prediction datasets at Nyx Farm, Prof Farm, and Probolinggo region. The three models are built to meet the needs of predictive information based on multidimensional data models, namely the time dimension, location dimension, and farm unit dimension, the location dimension is limited to the East Java+district area even though the location hierarchy is presented up to the village.

Table 4. LSTM model evaluation results [17]

Datasets	RMSE	MAPE	R^2
Prof farm	1.896	0.057	0.765
Nyx Farm	1.004	0.061	0.367
Probolinggo	1.620	0.039	0.782

3.2. Multidimensional model schema selection results for traditional BI of dairy goat dairy businesses

The results of the multidimensional model scheme selection of dairy goat milk BI are shown in Table 5. OLAP-1 and OLAP-2 testing resulted in the same visualization. Differences in results are seen in OLAP-3, OLAP-4, and OLAP-5 tests.

Table 5. Test results of the galaxy model scheme against the OLAP scenario

Test ID	Expected results	Status (True=1; False=0)	
		Galaxy Scenario 1	Galaxy Scenario 2
OLAP-3	Display information on the total production and distribution of each dairy product, namely dairy milk, pasteurized milk, kefir, yogurt, and cheese.	0	1
	Displaying a comparison of dairy milk production and distribution trends	0	1
	Displays a comparison of the amount of milk distribution for each partner	0	1
	Displays the percentage of delivery (dairy milk distribution)	0	1
	Displays the total revenue from the sale of dairy milk	0	1
	Displaying the average selling price of dairy milk	0	1
	Displays the lowest selling price of fresh milk	0	1
	Displays the highest selling price of fresh milk	0	1
	Displays the annual trend of dairy milk production	1	1
	Displays the monthly trend of dairy milk distribution	0	1
	Displays the percentage of production of each dairy product such as dairy milk, pasteurized milk, kefir, yogurt, and cheese	1	1
	Displays the percentage of distribution of each dairy product such as dairy milk, pasteurized milk, kefir, yogurt, and cheese	0	1
OLAP-4	Display information on the total production and distribution of each dairy product, namely dairy milk, pasteurized milk, kefir, yogurt and cheese.	0	1
	Displaying a comparison of dairy milk production and distribution trends	0	1
	Displays a comparison of the amount of milk distribution for each partner	0	1
	Displays the percentage of delivery (dairy milk distribution)	0	1
	Displays the total revenue from the sale of dairy milk	0	1
	Displaying the average selling price of dairy milk	0	1
	Displays the lowest selling price of fresh milk	0	1
	Displays the highest selling price of fresh milk	0	1
	Displays the annual trend of dairy milk production	1	1
	Displays the monthly trend of dairy milk distribution	0	1
	Displays the percentage of production of each dairy product such as dairy milk, pasteurized milk, kefir, yogurt, and cheese	1	1
	Displays the percentage of distribution of each dairy product such as dairy milk, pasteurized milk, kefir, yogurt, and cheese	0	1
OLAP-5	Display information on the total production and distribution of each dairy product, namely dairy milk, pasteurized milk, kefir, yogurt and cheese.	0	1
	Displaying a comparison of dairy milk production and distribution trends	0	1
	Displays a comparison of the amount of milk distribution for each partner	0	1
	Displays the percentage of delivery (dairy milk distribution)	0	1
	Displays the total revenue from the sale of dairy milk	0	1
	Displaying the average selling price of dairy milk	0	1
	Displays the lowest selling price of fresh milk	0	1
	Displays the highest selling price of fresh milk	0	1
	Displays the annual trend of dairy milk production	1	1
	Displays the monthly trend of dairy milk distribution	0	1
	Displays the percentage of production of each dairy product such as dairy milk, pasteurized milk, kefir, yogurt, and cheese	1	1
	Displays the percentage of distribution of each dairy product such as dairy milk, pasteurized milk, kefir, yogurt, and cheese	0	1

OLAP scenario 3 was tested by drilling down and dice the time dimension by selecting the time value of year=2023 and the dimension value of province=Jawa Timur, while the farm unit dimension was "all". Both scenarios display different results, the galaxy schema of scenario 2 displays all correct information as shown in Figure 10. The distribution information displayed by both scenarios has significant differences. Furthermore, validation is carried out on the galaxy scheme scenario 2 by filtering the *fact_distribution* data with *id_product_type*=5, namely the distribution of kefir-type milk as shown in Figure 10.

The location of Probolinggo, East Java has *id_location*=47,417 as shown in Figure 11. In Figure 12, kefir production is shown in only 4 rows with an *id_time* value greater than 365 and an *amount_distribution* value=3004. However, the number of distributions of 3004 is not only distribution from East Java but also other provincial areas. The results of OLAP 03 visualization in the galaxy scheme scenario 1 display incorrect information because the amount of distribution displayed is all regions not just East Java, namely 3004, while the results of scenario 2 visualization display correct information that the amount of distribution of kefir products in East Java in 2023 is 2 liters from Probolinggo region.

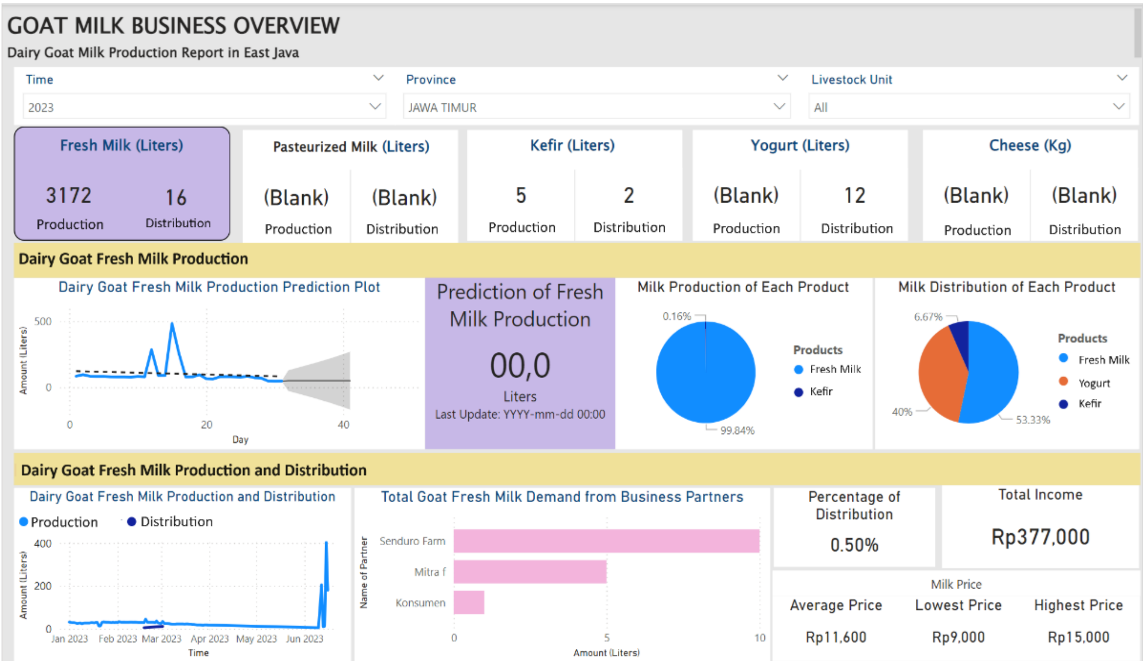


Figure 10. OLAP-03 Galaxy schema scenario 2 information display in BI dashboard – all true

id_time	id_location	id_livestock_unit	id_product_type	id_businesspatner	distribution_amount
529	27721	1	5	7	0
516	27721	1	5	7	3000
455	27760	4	5	12	2
414	47417	9	5	15	2

Figure 11. Validation of OLAP 03 results on distribution facts of Kefir product types

id	province	district/city	subdistrict	village	created_dt	modified_dt
47417	JAWA TIMUR	KAB. PROBOLINGGO	Tiris	Tlogoargo		

id	province	district/city	subdistrict	village	created_dt	modified_dt
27721	DKI JAKARTA	KOTA ADM. JAKARTA SELATAN	Pasar Minggu	Pasar Minggu		

Figure 12. Description of East Java locations in location dimensions

Based on the results of the OLAP scenario validation test on the two scenario schemes tested, the galaxy model scheme scenario 2 was selected as shown in Figure 6. The selected galaxy scheme explains that production and distribution facts are integrated with the dimensions of time, location, farm units, and product types, while the supply source dimension only has a relationship with production facts and the business partner dimension only has a relationship with distribution facts.

3.3. Real-time business intelligence development using machine learning

The development of BI to RTBI focuses on an integrated process between layers depicted through the RTBI architecture in Figure 4. The integrated process will start as soon as new data enters the RTDW, depicted with a real-time analysis sequence diagram as shown in Figure 13. There are 2 actors in the RTBI system, namely livestock units and stakeholders. The livestock unit inputs operational data, while stakeholders act as business actors who need business information for business improvement strategy needs. Livestock units can also be stakeholders. The livestock unit sends new data to the online application and historical data is added to the RTDW. The new data entry event suppresses services in the WebSocket path. The system was built using the Kafka service as a streaming data pipeline and WebSocket as a protocol that stores analytical modules. The analysis results are then saved into RTDW and WebSocket to be displayed in the RTBI dashboard. The DW scheme was adopted into RTDW, however, there was a development of a fact table for real-time system needs, namely an incremental milk prediction fact table. The milk prediction fact table was built to store prediction results and historical data replicas of milk production facts for model retraining. OLAP and ML models are entered using the application programming interface (API) module. The API contains translations of OLAP and ML model execution via the WebSocket path. WebSocket is a two-way real-time communication path between client and server so that the resulting latency is lower than one-way web protocols.

The RTBI dashboard is designed based on the traditional BI dashboard template. The dashboard design is built on a website basis. All integrated processes of the RTBI framework are implemented and the final information visualization is displayed in the RTBI dashboard prototype as shown in Figure 14. Information on the prediction plot for fresh milk production for dairy goats is taken from the production facts table as actual data and *pred_susu* facts as predicted data. The latest predicted figures are also presented in the RTBI dashboard.

3.4. Result measurement of analysis latency RTBI application

Latency results from real-time milk prediction simulations carried out on Nyx Farm and Prof Farm farms can be seen in Figure 15. The latency produced in 14 simulations shows that the prediction speed is quite stable starting from the 2nd simulation. Real-time analysis latency in Nyx Farm farming units has an average of 9.85×10-5 ms as shown in Figure 15(a), while Prof Farm has an average latency of 9.35×10^{-5} ms as shown in Figure 15(b).

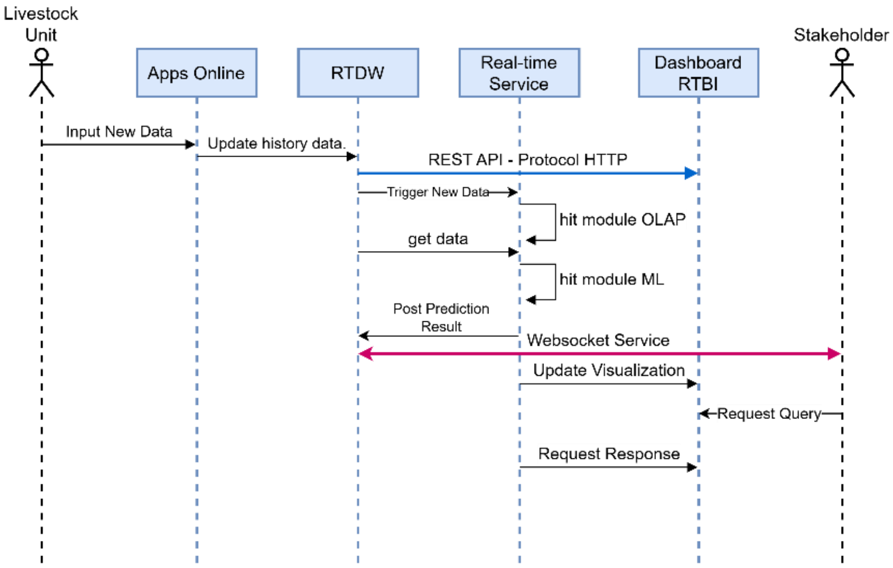


Figure 13. Sequence diagram real-time analysis in RTBI

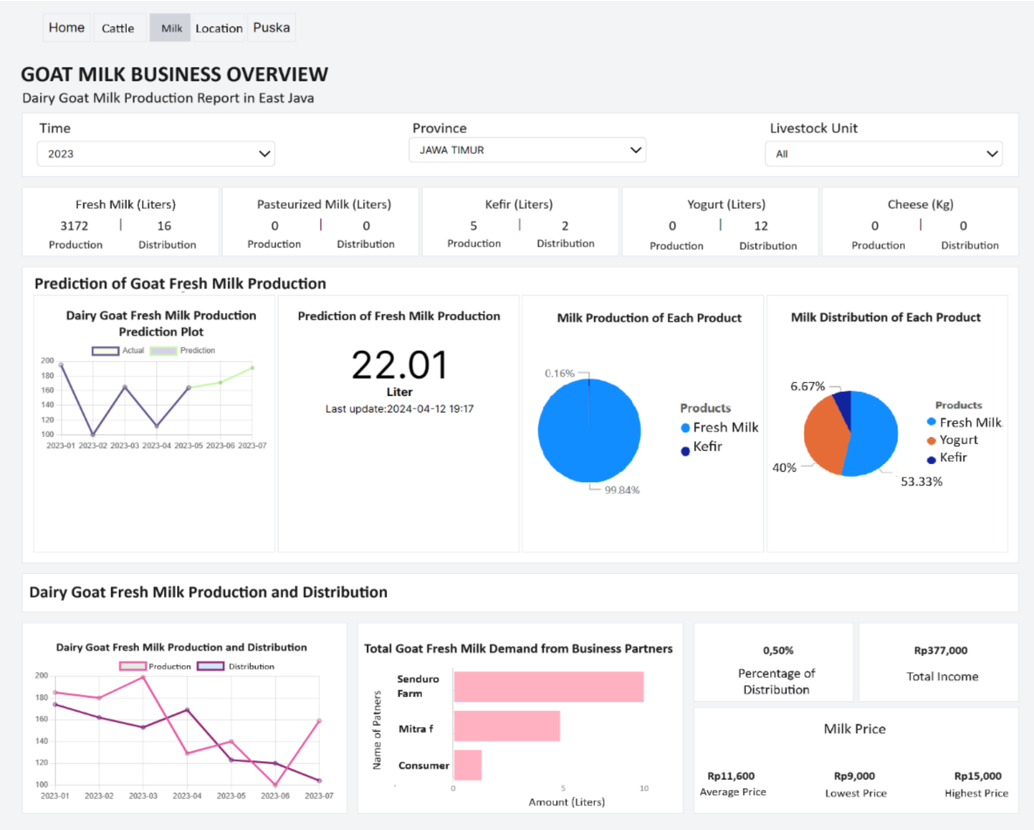
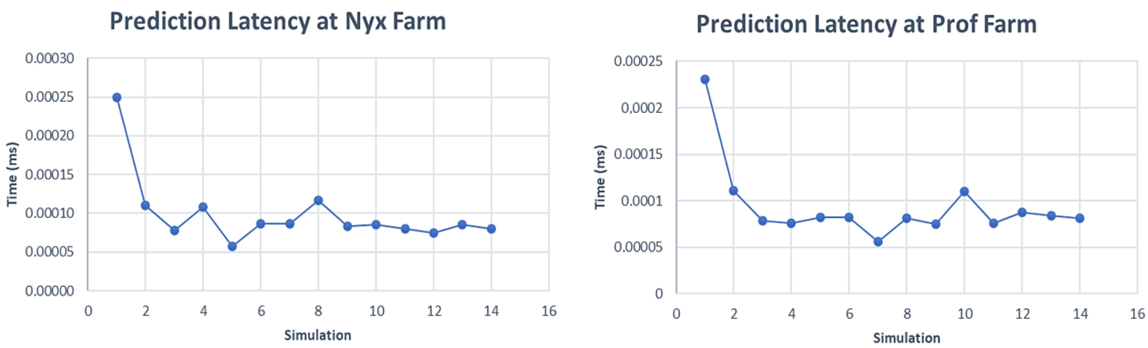


Figure 14. Real-time BI dashboard prototype overview



(a) (b)

Figure 15. Real-time prediction analysis latency plot for 14 daily prediction simulations (a) Nyx Farm and (b) Prof Farm

4. CONCLUSION

RTBI with a ML approach was successfully built from scratch in a case study of the dairy goat milk business in Probolinggo, East Java, which does not yet have business intelligence technology with the smallest average analysis latency approaching 0 milliseconds, namely 9. 35×10^{-5} ms. The integration of each layer in the developed RTBI framework was successfully designed into the RTBI architecture. The implementation of the RTBI architecture starts from the design of the traditional BI conceptual model. The success of the RTBI project in the dairy goat milk business case study was due to the correct traditional BI conceptual model being built first. The output of the traditional BI conceptual model is in the form of the best multidimensional model schema which is centered on data for descriptive analysis using OLAP and predictive using ML. The galaxy model scheme for scenario 2 is the best multidimensional model scheme and the LSTM model is the best ML model integrated into the RTBI framework. The integrated process for each layer was also successfully carried out through a real-time system integration layer using a bi-directional connection protocol as a communication path between client-server such as WebSocket. The role of Kafka and Docker data streams as containers that hold all real-time services is also an important tool in building integrated processes between layers in the RTBI framework. The output of this research conceptual model will contribute to developing real-time data warehouse technology for goat farming in East Java. However, real data is minimal, so RTBI is less able to depict the movement of the business.

Based on the results of this research, it is possible to implement the RTBI framework for general use. However, the multidimensional data and machine learning models used need to be rebuilt to suit business information needs and data characteristics. The LSTM model used in the RTBI dairy goat milk business can also be developed in the future to increase accuracy while maintaining low analysis latency.

ACKNOWLEDGEMENTS

We would like to thank the Department of Animal Husbandry and the goat farming unit in Probolinggo, East Java, especially Prof Farm and Nyx Farm who have been willing to provide documentation of goat milk production data for 2022 to 2023.

REFERENCES

[1] M. Rath, "Realization of business intelligence using machine learning," *Internet of Things in Business Transformation: Developing an Engineering and Business Strategy for Industry 5.0*, pp. 169–184, 2021, doi: 10.1002/9781119711148.ch10.
[2] A. K. Mohamad, M. Jayakrishnan, and M. Mohd Yusof, "Thriving information system through business intelligence knowledge management excellence framework," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 1, pp. 506–514, Feb. 2022, doi: 10.11591/ijece.v12i1.pp506-514.
[3] A. Vaisman and E. Zimányi, "Data warehouse systems: design and implementation," *Data Warehouse Systems: Design and Implementation*, pp. 1–625, 2014, doi: 10.1007/978-3-642-54655-6.
[4] A. Dhaouadi, K. Bousselmi, M. M. Gammoudi, S. Monnet, and S. Hammoudi, "Data warehousing process modeling from classical approaches to new trends: Main features and comparisons," *Data*, vol. 7, no. 8, p. 113, Aug. 2022, doi: 10.3390/data7080113.
[5] R. Esbai, S. Hakkou, and M. A. Habri, "Modeling and automatic generation of data warehouse using model-driven transformation in business intelligence process," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 3, pp. 1866–1874, Jun. 2023, doi: 10.11591/ijeecs.v30.i3.pp1866-1874.
[6] G. V. Machado, Í. Cunha, A. C. M. Pereira, and L. B. Oliveira, "DOD-ETL: distributed on-demand ETL for near real-time business intelligence," *Journal of Internet Services and Applications*, vol. 10, no. 1, Nov. 2019, doi: 10.1186/s13174-019-0121-z.
[7] S. Fong and Y. Hang, "Enabling real-time business intelligence by stream data mining," in *New Fundamental Technologies in Data Mining*, InTech, 2011. doi: 10.5772/13123.
[8] A. P. S. Pall and J. Singh, "ETL methodologies, limitations and framework for the selection and development of an ETL tool," *International Journal of Research in Engineering and Applied Sciences (IJREAS)*, vol. 6, no. 5, pp. 108–114, 2016.
[9] F. de Assis Vilela, V. C. Times, A. C. de Campos Bernardi, A. de Paula Freitas, and R. R. Ciferri, "A non-intrusive and reactive architecture to support real-time ETL processes in data warehousing environments," *Heliyon*, vol. 9, no. 5, May 2023, doi: 10.1016/j.heliyon.2023.e15728.
[10] S. YiChuan and X. Yao, "Research of real-time data warehouse storage strategy based on multi-level caches," *Physics Procedia*, vol. 25, pp. 2315–2321, 2012, doi: 10.1016/j.phpro.2012.03.390.
[11] J. Bharadiya and J. P. Bharadiya, "Machine learning and AI in business intelligence: Trends and opportunities," *International Journal of Computer (IJC)*, vol. 48, no. 1, pp. 123–134, 2023.
[12] R. Sharda, D. Delen, and E. Turban, *Business intelligence, analytics, and data science: a managerial perspective*. Pearson, 2018.
[13] Y. Ning, H. Kazemi, and P. Tahmasebi, "A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and prophet," *Computers and Geosciences*, vol. 164, Jul. 2022, doi: 10.1016/j.cageo.2022.105126.
[14] D. Borges and M. C. V Nascimento, "COVID-19 ICU demand forecasting: a two-stage Prophet-LSTM approach," *Applied Soft Computing*, vol. 125, Aug. 2022, doi: 10.1016/j.asoc.2022.109181.
[15] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM -- a tutorial into long short-term memory recurrent neural networks," *arXiv:1909.09586*, Sep. 2019.
[16] S. Arslan, "A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data," *PeerJ Computer Science*, vol. 8, Jun. 2022, doi: 10.7717/peerj-cs.1001.
[17] A. Primawati, I. S. Sitanggang, Annisa, and D. A. Astuti, "Performance LSTM and prophet for prediction time series with limited data: Case study of daily goat milk production," Dec. 2023. doi: 10.1109/icon-sonics59898.2023.10435067.
[18] İ. Kırbaş, A. Sözen, A. D. Tuncer, and F. Ş. Kazancıoğlu, "Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches," *Chaos, Solitons & Fractals*, vol. 138, Sep. 2020, doi: 10.1016/j.chaos.2020.110015.
[19] A. Primawati, I. S. Sukaesih, Annisa, and D. A. Astuti, "Performance comparison of LSTM and prophet for time series prediction (case study of daily cow milk production)," (in Bahasa), *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 9, no. 3, pp. 428–435, 2023, doi: 10.26418/jp.v9i3.72031.
[20] N. H. A. Rahman, M. Z. Hussin, S. I. Sulaiman, M. A. Hairuddin, and E. H. M. Saat, "Univariate and multivariate short-term solar power forecasting of 25MWac Pasir Gudang utility-scale photovoltaic system using LSTM approach," *Energy Reports*, vol. 9, pp. 387–393, Oct. 2023, doi: 10.1016/j.egyr.2023.09.018.
[21] M. A. Tripathi, K. Madhavi, V. S. P. Kandi, V. K. Nassa, B. Mallik, and M. K. Chakravarthi, "Machine learning models for evaluating the benefits of business intelligence systems," *The Journal of High Technology Management Research*, vol. 34, no. 2, Nov. 2023, doi: 10.1016/j.hitech.2023.100470.
[22] S. Bouaziz, A. Nabli, and F. Gargouri, "Design a data warehouse schema from document-oriented database," *Procedia Computer Science*, vol. 159, pp. 221–230, 2019, doi: 10.1016/j.procs.2019.09.177.

[23] H. K. Yaseen and A. M. Obaid, “Big data: definition, architecture and applications,” *JOIV: International Journal on Informatics Visualization*, vol. 4, no. 1, pp. 45–51, Feb. 2020, doi: 10.30630/joiv.4.1.292.

[24] L. Angel Valenzuela Ygnacio, M. Giraldo Retuerto, and L. Andrade-Arenas, “Mobile application with business intelligence to optimize the control process of tourist agencies,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 3, p. 1708, Mar. 2023, doi: 10.11591/ijeecs.v29.i3.pp1708-1718.

[25] J. E. Plazas *et al.*, “A conceptual data model and its automatic implementation for IoT-based business intelligence applications,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10719–10732, Oct. 2020, doi: 10.1109/jiot.2020.3016608.

[26] A. Primawati, I. S. Sitanggang, Annisa, and D. A. Astuti, “Business intelligence and analytics in dairy goat livestock: Current and future challenge,” *2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia, 2021, pp. 323-329, doi: 10.1109/ICIMCIS53775.2021.9699336.

[27] H. M. Merdas and A. H. Mousa, “Food sales prediction model using machine learning techniques,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6578–6585, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6578-6585.

[28] B. T. Khoa and T. T. Huynh, “Forecasting stock price movement direction by machine learning algorithm,” *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, pp. 6625–6634, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6625-6634.

[29] N. Biswas, S. Chattopadhyay, G. Mahapatra, S. Chatterjee, and K. C. Mondal, “SysML based conceptual ETL process modeling,” in *Computational Intelligence, Communications, and Business Analytics*, Springer Singapore, 2017, pp. 242–255. doi: 10.1007/978-981-10-6430-2_19.

BIOGRAPHIES OF AUTHORS

Alusyanti Primawati completed a bachelor's degree in informatics engineering from Indraprasta University PGRI, a master's degree in computer science from STMIK Nusa Mandiri and is currently pursuing doctoral education in computer science at IPB University. Currently working as a professional lecturer in informatics engineering at Indraprasta University PGRI with a scientific focus on knowledge-based systems, artificial neural networks and business intelligence. Her research interests include data science, machine learning, and business intelligence. She can be contacted at email: 1149alusyanti@apps.ipb.ac.id.

Imas Sukaesih Sitanggang received a Ph.D. degree in computer science from the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, in 2013. She is a lecturer in Computer Science Department, IPB University, Indonesia. Her main research interests include spatial data mining and smart agriculture. She can be contacted at email: imas.sitanggang@apps.ipb.ac.id.

Annisa is a doctorate in the Computer Science Department, IPB University. She is a doctoral graduate from Hiroshima University, Japan. Previously she earned a master's degree in computer science from the University of Indonesia. She focuses on the fields of data mining, skyline queries and data management. She can be contacted at email: Annisa@apps.ipb.ac.id.

Dewi Apri Astuti is a professor at the Faculty of Animal Husbandry, Bogor Agricultural University. He conducted research in the field of Animal Nutrition. She graduated from the UGM Faculty of Animal Husbandry in 1984. The field of animal husbandry has become his scientific focus, one of which is ruminants. She can be contacted at email: dewiapriastuti86@gmail.com.