CoreID JOURNAL

Vol. 3 No. 2, July 2025 (65-78)
ISSN: 2987-6990
DOI: https://doi.org/10.60005/coreid.v3i2.136

Academic Data Quality Measurement in SALAM Application Using Six Sigma Method

Imam Firdaus[1], Cecep Nurul Alam[2], Yana Aditia Gerhana[3], Mohamad Irfan[4], Ibrahim Iskandar[5]
[1,2,3,4] Informatics Department, UIN Sunan Gunung Djati Bandung, Indonesia
[5]IAIN Kendari, Indonesia

Article Info

Article history:

Received April 07, 2025
Revised May 07, 2025
Accepted July 17, 2025

Keywords:

Measurement of data quality
DMAIC
six sigma

ABSTRACT

Data quality plays a critical role in ensuring the reliability and usefulness of information for decision making in higher education institutions. However, academic data within the SALAM application at UIN Sunan Gunung Djati Bandung has not previously undergone a systematic quality assessment, leading to uncertainty in several managerial and academic decisions. This study aims to evaluate the quality of academic data in the SALAM application using the Six Sigma method with the DMAIC (Define–Measure–Analyze–Improve–Control) framework. Five data quality dimensions completeness, consistency, conformity, uniqueness, and timeliness are employed to measure and analyze data quality performance. The measurement process begins with data definition and extraction, followed by quantitative analysis using sigma metrics. The results indicate that the overall quality of academic data is at a moderate level, with an average sigma score of approximately 3, primarily influenced by incomplete and inconsistent data. In contrast, the timeliness dimension demonstrates excellent performance, achieving a sigma metric of 6 due to the long-term availability of data over more than ten years. This study contributes by providing an empirical, data-driven evaluation of academic data quality and offers practical insights for implementing continuous monitoring and improvement strategies to enhance data reliability and support more effective decision making in higher education institutions.

Corresponding Author:

Imam Firdaus
UIN Sunan Gunung Djati Bandung, Cibiru, Bandung, West Java, Indonesia.
Email: imamfirdaus@uinsgd.ac.id

1. INTRODUCTION

Etymologically, data is defined as the plural form of datum which in Latin means a statement or value of a reality. The statement or value comes from the process of measuring or observing a variable, and is represented in the singular or plural form of numbers, characters, images, or sounds. Data can be in the form of files and strings which are generally divided into two types, namely confidential and non-confidential data [1].

A collection of data or facts that have been processed and managed properly so that they become something that is easy to understand and useful for the recipient will become information. However, not all data can be processed into information that is useful for the recipient. This is because if the data has poor quality, the information produced will also be bad for the recipient [2].

Thus, it is important to ensure that the data used is of good quality in order to produce useful and reliable information. Poor data quality can lead to errors, inaccuracies, and misinterpretations in decision making and analysis. Therefore, maintaining data quality is an important factor in ensuring that the information produced is useful and reliable for the recipient [3].

Figure 1. Data quality pyramid

Data quality measurement can be seen through the data quality dimensions as depicted in Figure 1.1. Data quality assessment has been discussed intensively in practice and research. Data quality measurement is important to support quality management and economically oriented decision making under uncertainty, it is important to assess the level of data quality using predetermined metrics [4]

The Academic Service Administration System (SALAM) is a crucial application owned by UIN Sunan Gunung Djati Bandung. This application is the main application used by the academic community of UIN Sunan Gunung Djati Bandung in the lecture process. The presence of the SALAM application is a hope to improve the quality of graduates and the Marwah of the campus according to what was conveyed by the Chancellor of UIN Sunan Gunung Djati Bandung for the 2019-2023 period, Prof. Mahmud at the launch of the New SALAM application at the 2021 academic coordination meeting [5].

Data processing and academic services in higher education institutions that are carried out optimally and integrated are very much needed to realize excellent service and build a positive image in the campus environment. However, this concept ultimately raises doubts because the quality of the data contained in the SALAM application is not yet known. If the quality of the data is poor, it can cause negative impacts such as reputational risks, violations of data protection regulations, negative media coverage, missed opportunities, and failure to provide the best service to users.

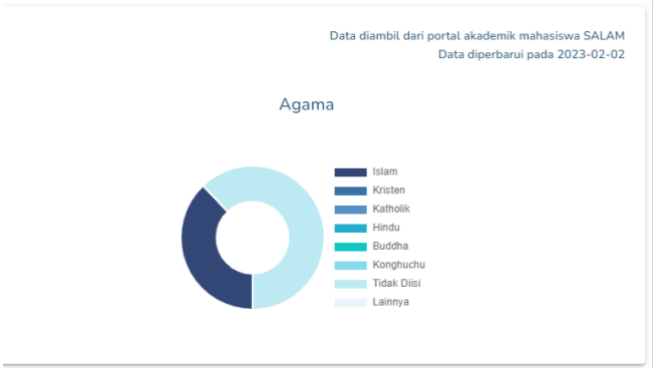

Figure 2. Statistics of religious data on the TERAS application sourced from the SALAM application

A statistic of student religious data on the TERAS application of UIN Sunan Gunung Djati Bandung which is sourced from the SALAM application. The TERAS application is an Executive Dashboard used by the leadership of UIN Sunan Gunung Djati Bandung. The statistical data shows that more than 50% of students do not complete religious information on the SALAM application.

Academic data owned by UIN Sunan Gunung Djati Bandung is included in the big data category because of its large amount. In the context of big data, poor data quality is a crucial problem because there are many characteristics of big data that have a direct impact on data quality [7]. The quality of the data owned has an important role in ensuring that the decision will have a positive impact, conversely if the quality of the data owned is poor, it can lead to decisions that have a negative impact on the performance of the managed system and the management of available resources [6].

Didem Gürdüra, Jad El-khourya, Mattias Nybergb in 2018 in their journal explained about the Methodology for linked enterprise data quality assessment through information visualizations. The results of the study indicate that visualized data quality measurements can show overall information to identify patterns to better understand the situation to realize the relationship between resources in big data processing [7].

One of the big data management can apply the TQM concept. Total Quality Management (TQM) is a management approach that focuses on continuous efforts to improve the quality of products, services, and organizational processes as a whole [8]. The TQM concept emphasizes the importance of involving all members of the organization, from management to employees, in an effort to achieve superior quality and higher customer satisfaction. The main principles of TQM include customer orientation, strong leadership, employee participation, a fact-based approach to decision making, and continuous improvement [9].

One of the methodologies used in the implementation of TQM is Six Sigma [10]. In 2019 Siim Koppel, Shing Chang in his journal MDAIC - a Six sigma implementation strategy in big data environments explained that Six sigma was proposed as a manufacturing discipline because data can be well controlled so that problem solving can be identified according to the data [11]. The six sigma method is not only used by the manufacturing sector, but is also used to measure data quality in the service sector [12], so the Six sigma method can be used to measure the quality of academic data in the SALAM application. Six sigma is a statistical-based quality improvement method that requires high discipline and is carried out comprehensively to identify and address the main sources of problems with the DMAIC (Define-Measure-Analyze-Improve-Control) approach [13].

The Six sigma method is a structured approach to improving processes, with a focus on efforts to reduce process variation and reduce data errors using statistical tools and intensive problem solving. The Six sigma method is known as an effective method in improving data quality, with a target defect of no more than 3.4 per 1 million opportunities [14]

Measuring the quality of academic data in the SALAM application provides a number of benefits for UIN Sunan Gunung Djati Bandung, including:

1. Improving the quality of decision making. By having accurate and complete academic data, UIN Sunan Gunung Djati Bandung can make better decisions regarding future academic planning.
2. Improve the efficiency and effectiveness of the academic system. Accurate and up-to-date academic data can help UIN Sunan Gunung Djati Bandung in monitoring student progress, evaluating curriculum and study programs, and planning resource and budget allocation more efficiently.
3. Increase public trust. By having accurate and reliable academic data, UIN Sunan Gunung Djati Bandung can increase public trust, to help promote its good reputation.

Based on various studies that have been presented, a study was conducted on "Measurement of Academic Data Quality in the SALAM Application Using the Six Sigma Method".

2. METHOD

2.1 DMAIC model for Six sigma method

DMAIC (Define, Measure, Analyze, Improve, Control) is a model used in the Six Sigma method to solve problems and improve the quality of the process from data. The DMAIC model consists of five steps that are executed sequentially to achieve significant improvements in a process.

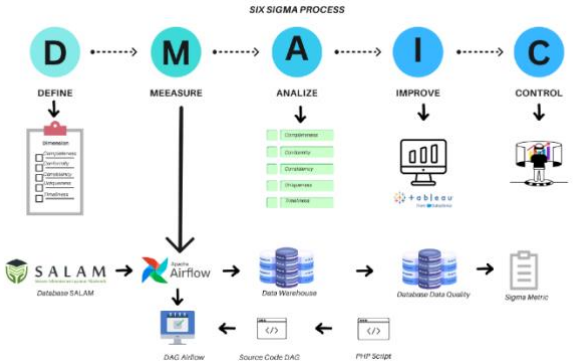

Figure 3. DMAIC model process for Six sigma method

The data quality measurement process begins with defining the problem or opportunity for improvement clearly and specifically. The purpose of this step is to understand the final goal to be achieved from the data quality measurement process carried out. This step will also explain what attributes will be measured for data quality and the dimensions of the data quality measurement.

The next step is to collect and measure data quality. In this step, data is drawn to then measure data quality by calculating DPMO (Defects-Per-Million-Opportunities) from each attribute in each dimension to obtain its sigma metrics score. This data will provide an understanding of the extent to which the current process is running and provide insight into the root causes of the problems faced.

After the measurement is carried out, the next step is to analyze the results of the measurements that have been carried out to identify the root causes of the data quality problems. The measurement data from each dimension will be displayed and the average of each attribute will be calculated.

After the root cause of the problem is identified, the next step is to find solutions and implement the changes needed to improve data quality. The measurement results will be visualized in graphical form to make it easier to find solutions based on the attributes that are the main causes of the data quality score.

The final step in the DMAIC model is to check and sustain the improvements that have been made. To achieve this, a control system is implemented to continuously monitor the performance of the process after the improvements have been made. Relevant quality metrics are continuously monitored, and if there is any non-conformance with the established targets, corrective actions are taken. The goal of this step is to ensure the sustainability of the improvements in the long term and prevent the same problems from recurring.

2.2 Define

This section explains the measurement process that will be carried out. The define stage in the Six sigma method is used to identify the formulation of problems in the quality of academic data contained in the SALAM UIN Sunan Gunung Djati Bandung application.

The purpose of this study, as explained in the previous chapter, is to determine the level of quality of academic data contained in the SALAM UIN Sunan Gunung Djati Bandung application. The quality of academic data in the SALAM UIN Sunan Gunung Djati Bandung application will be measured using the Six sigma method with 5 dimensions of measurement, namely completeness, conformity, uniqueness, consistency, and timeliness.

The SALAM application is a very important application for the academic community of UIN Sunan Gunung Djati Bandung, considering that this application is an academic service administration system application to support daily lecture activities. So that academic data owned by students, lecturers, and employees will be centered on this application.

Data will be measured in quality if it has CTQ (Critical To Quality). CTQ is very important in measuring data quality because it will help us identify the attributes of the data that are important enough to meet the needs or expectations of data use. CTQ can help us identify key parameters that must be measured and evaluated to ensure that the data produced is of high quality. For example, CTQ can include attributes such as data accuracy, data integrity, data availability, and so on.

Data quality can be defined based on its properties which are often referred to as metrics or data quality dimensions. The data quality measurement dimension functions to map dimensions to values between 0 and 1 and measure aspects and parts of data quality related to this quality dimension [50]. In order to be able to design more focused and effective data quality measurements, and also to be able to take appropriate and accurate corrective actions to improve the quality of academic data in the SALAM application, the following are the dimensions that are the standard in measuring the quality of academic data in the SALAM application.

- Completeness, The measurement standard carried out on the completeness dimension is based on the amount of data filled in from each academic data attribute in the SALAM application.
- Conformity, The measurement standard carried out on the uniqueness dimension is based on the number of unique or non-duplicate data on the academic data attribute in the SALAM application.
- Uniqueness, The measurement standard carried out on the consistency dimension is based on the number of data in terms of the consistency of data filled in according to the related attributes in the SALAM application.
- Consistency, The measurement standard carried out on the conformity dimension is based on the number of data type conformities in the filled data in the academic data attribute of the SALAM application.
- Timeliness, The measurement standard carried out on the Timeliness dimension is based on the consistency of the availability of academic data in the SALAM application from year to year.

2.3 Measure

The measure process is the next stage after defining in the initial stage of research findings. This process is a stage for collecting data from database sources, namely the SALAM application database to the quality database in the data warehouse.

In this measure process, it is explained how data from the SALAM application can be pulled using apache airflow to then enter the data warehouse, precisely in the quality database. In the database, the results of the DPMO measurement and the sigma metric level of the attributes that are measured for data quality will appear.

This section will explain how the program code structure of airflow uses the PHP programming language and the DAG program code in the python programming language to be able to pull data from the SALAM application. Each dimension of data quality measurement will explain how the query is needed to pull data into the data warehouse from the SALAM application.

2.4 Analyze

The analyze process is the last process in research findings before entering the improve stage in the discussion of research results. This process is a stage to identify the root cause of the problem which will then be evaluated data from the results of measuring the quality of academic data in the SALAM application. This process will collect data types, total data, total defective data, DPMO values, and sigma metric scores from each dimension whose data quality is measured.

After the data retrieval process using apache airflow, the results of measuring the quality of academic data in the SALAM application will be stored in each table according to the table name created in the apache airflow DAG. Each data attribute will be stored according to the dimensions taken in the process of measuring the quality of academic data in the SALAM application. DPMO and sigma metric scores will appear in each table.

The data that has been stored is then processed into average data for each dimension. From the results of this processing, the average value of each dimension will be obtained which is then averaged back to get the final score for the level of academic data quality in the SALAM application. These data will be the main capital for the discussion section of the research results, especially for visualizing data in the improve section.

2.5 Improve

The improve section is the first part of the discussion of the research results on the results and discussion. This section focuses on the implementation and development of improvement solutions that can be proposed to improve the quality of academic data at UIN Sunan Gunung Djati Bandung contained in the SALAM application. This section displays data visualization from the results of measuring the quality of academic data in the SALAM application.

The results of measuring the quality of academic data in the SALAM application will later be visualized in the form of graphs using the Tableau application. The visualized data includes the results of DPMO measurements and sigma metric scores of the attributes according to their measurement dimensions. The average results of data quality measurements will be visualized and sorted from dimensions that greatly affect poor data quality.

In addition to the results of the overall average calculation, each dimension will also be visualized based on its DPMO and sigma metric. Each data in the dimension will be sorted from its influence in affecting poor data quality. Data that has been sorted by influence will be the initial capital in the control section to be a reference for improving data quality.

2.6 Control

The control section is the last section of the discussion of research results on the results and discussion. This section focuses on the aim of ensuring that the improvements that have been made are maintained and provide sustainable results. The process will involve the development and implementation of controls that must ensure that the improved process continues to run consistently. This control section is also expected to provide results according to expectations.

The focus of control carried out on measuring the quality of academic data in the SALAM application starts from the Develop Control Plan. The development of this control plan aims to control the quality value of academic data owned by UIN Sunan Gunung Djati Bandung in a better direction.

Furthermore, the development of a monitoring system or Establish Monitoring Systems is carried out. Currently, UIN Sunan Gunung Djati Bandung already has a monitoring system to visualize data owned by UIN Sunan Gunung Djati Bandung as an executive dashboard for leaders, namely TERAS UIN Sunan Gunung Djati Bandung. The results of measuring the quality of academic data can later be entered into the application as a data quality module.

The next process is Implement Process Controls. This process is a fairly important step in the process of improving the quality of academic data. Providing periodic warnings to the academic community of UIN Sunan Gunung Djati Bandung to update and complete the data contained in the SALAM application is also important so that the data contained in the SALAM application is the latest data. The last step in this control process is Continuous Monitoring and Improvement. Continuous monitoring and improvement must be carried out periodically every day. This monitoring and improvement will be scheduled every day to update data by periodically pulling data using apache airflow from the SALAM application database to data quality in the data warehouse.

3. RESULTS AND DISCUSSION

3.1 Define

The Define process is the first step in the DMAIC method in SIX SIGMA. This step involves identifying and clearly defining the project objectives. In the context of measuring academic data quality, the use of relevant standards is very important. One of the standards used is PERMENRISTEKDIKTI Number 61 of 2016 concerning the Higher Education Database. This standard provides guidelines on how academic data should be measured and evaluated.

Based on DAMA DMBoK 2nd Edition, there are several dimensions that are standards in measuring data quality. In measuring the quality of academic data in the SALAM application, measurements are made based on the following five dimensions.

Table 1. Table Type Styles

No.	Dimension	Measurement Standards
1.	*Completeness*	The measurement standard carried out on the completeness dimension is based on the amount of data filled in from each academic data attribute in the SALAM application.
2.	*Uniqueness*	The measurement standard carried out on the uniqueness dimension is based on the amount of unique or non-duplicate data on the academic data attribute in the SALAM application.
3.	*Consistency*	The measurement standard carried out on the consistency dimension is based on the amount of data in terms of the consistency of the amount of data filled in according to the related attribute in the SALAM application.
4.	*Conformity*	The measurement standard carried out on the conformity dimension is based on the number of data type conformities in the filled data in the academic data attributes of the SALAM application. The measurement standard carried out on the Timeliness dimension is based on the consistency of the availability of academic data in the SALAM application from year to year.
5.	*Timeliness*	The measurement standard carried out on the conformity dimension is based on the number of data type conformities in the filled data in the academic data attributes of the SALAM application. The measurement standard carried out on the Timeliness dimension is based on the consistency of the availability of academic data in the SALAM application from year to year.

In measuring the quality of academic data in the SALAM application, the first measurement is carried out on the completeness dimension. This dimension examines the extent to which the data covers all required attributes. To measure completeness, calculations are carried out to determine the amount of data filled in from each existing attribute. The higher the percentage of data, the better the quality.

The measurement of the completeness dimension is based on the amount of data filled in from each attribute. For example, there are columns for Student ID, Student ID, and full name. These columns must be filled with data according to their attributes and must not be empty or NULL.

The next measurement is carried out on the uniqueness dimension which is an important factor in measuring data quality. This dimension assesses the extent to which data is unique or not duplicated. Measurement is carried out by calculating the total unique data from the entire data set. The higher the number of unique data, the higher the data quality in terms of uniqueness.

The measurement of the uniqueness dimension is based on unique data on each attribute whose data quality is checked. For example, there are student IDs, NIMs, and NIKs where each student can be sure to have different data, unless they continue their studies back at UIN Sunan Gunung Djati Bandung, then there is a possibility of duplicate student NIK data. However, in some of these cases, there are exceptions to the queries that are carried out.

Consistency is the next dimension that is evaluated in measuring the quality of academic data in the SALAM application. In this dimension, calculations are carried out to measure the consistency of the data filled in for each attribute. Consistent data indicates uniformity in filling in the same attribute.

Measurements on the consistency dimension are carried out by measuring the consistency of academic data from each attribute that is measured for data quality. For example, there are attributes of Student ID, NIM, and full name that have different standards for their consistency. Student ID has a standard that must be filled with numeric data, then NIM data must consist of numbers and a minimum of 10 digits, while the full name attribute must consist of letters and/or contain characters and must not contain numbers in it.

The conformity dimension is also the focus of measuring the quality of academic data in the SALAM application. This dimension evaluates the extent to which the data matches the data type set for each attribute. Measurements are made by comparing the expected data type with the actual data type in the data set. The higher the level of conformity, the better the data quality in the conformity dimension.

In the conformity measurement standard, measurements are carried out with the standard that data must be filled in and in accordance with the data type of the attribute.

Finally, the timeliness dimension evaluates the availability of academic data from year to year. Measurement is done by looking at the extent to which academic data is available in the SALAM application. By monitoring the availability of data from year to year, it can be seen whether the academic data is current and relevant. The timeliness dimension is very important to ensure that the data used in analysis or decision making for leaders is up-to-date. Measure

The measure process is a process of measuring the quality of data from the define process that has been done previously. This process involves collecting data from the SALAM application database source that will be pulled into the warehouse database which will then undergo a data quality measurement process. The required data will go through a data retrieval process with a scheme as shown in the following image.

Figure 4. Data retrieval scheme from SALAM application database

The first process carried out in the process of extracting data from the SALAM application is to use apache airflow as an open source workflow management platform for the extract-transform-and-load (ETL) process to the data warehouse. In this process, a program code is created as a function to extract the data needed from the SALAM application to measure its data quality. Each dimension has a different program code depending on the measurement standard used. The program code is created using the PHP programming language with a code structure as described in the following image.

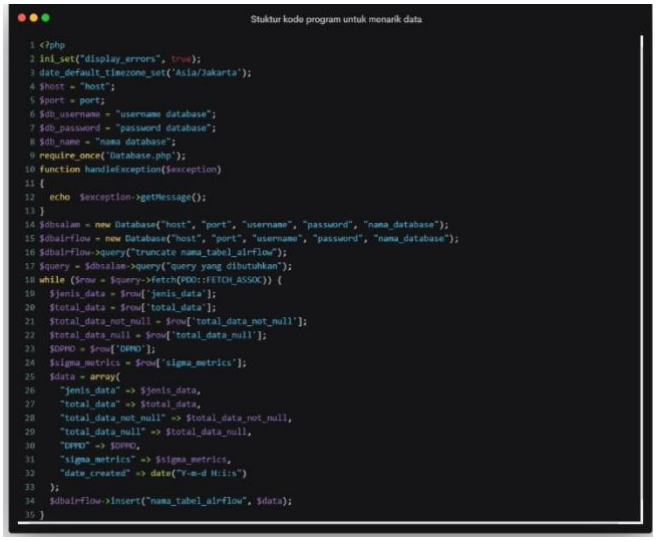

Figure 5. Program code structure to pull data from SALAM application using apache airflow to datamart database

The program code functions to pull data from the SALAM application database to the datamart database. The first line of the program code explains the service that will be run by mentioning the data source, data destination, and code writing. The next line explains the option settings to display PHP errors ("display_errors") so that you can see error messages if an error occurs in the data retrieval process. The time zone is set to "Asia/Jakarta" by default to perform scheduled data retrieval.

The variables that will be used to connect to the database are defined first and filled with the values with the appropriate host, port, username, password, and database name. Then the database file is imported. PHP which contains database classes to manage database connections and operations. The "handleException()" function is used to handle exceptions that occur if exceptions are needed depending on the dimensions for which the data quality measurement is carried out.

Objects for the SALAM and airflow databases are created with the appropriate host, port, username, password, and database name parameters. To avoid data repetition, data is deleted from the destination table before data retrieval is carried out. Then the required query is written according to the data retrieval needs.

Each row of query results uses the fetch(PDO::FETCH_ASSOC) method on the $query object. Then, the data on each row is stored in the appropriate variables, such as $data_type, $total_data, etc. The data is then inserted into the $data array along with the current date and time. Finally, the data in the $data array is inserted into the "airflow_table_name" table using the Insert() method on the $dbairflow object.

The academic data pulled into the datamart is the total data from the data from each attribute and the total defect data. The defect parameters of a data vary depending on the dimension of the academic data quality measurement taken. From the total defect data and the total overall data, the results of the DPMO (Defect per Million Objects) calculation are then obtained which determines the sigma metric value of each data attribute.

Potongan kode program airflow untuk menarik data dari basis data aplikasi SALAM ke datamart pada dimensi completeness

```sql
SELECT jenis_data,
SUM(total_data) AS total_data,
SUM(total_data_not_null) AS total_data_not_null,
SUM(total_data_null) AS total_data_null,
DPMO,
CASE
    WHEN DPMO <= 0 THEN 0
    WHEN DPMO <= 3.4 THEN 6
    WHEN DPMO <= 5 THEN 5.9
    WHEN DPMO <= 9 THEN 5.8
    WHEN DPMO <= 12 THEN 5.7
    -- tutorial DPMO lainnya --
    ELSE 0
END AS sigma_metrics
FROM (
SELECT jenis_data,
    COUNT(mhs_id) AS total_data,
    SUM(CASE WHEN data IS NOT NULL THEN 1 ELSE 0 END) AS total_data_not_null,
    SUM(CASE WHEN data IS NULL THEN 1 ELSE 0 END) AS total_data_null,
    (SUM(CASE WHEN data IS NULL THEN 1 ELSE 0 END) / COUNT(mhs_id)) * 1000000 AS DPMO
FROM (
  SELECT 'ID Mahasiswa' AS jenis_data, mhs_id AS data, mhs_id
     FROM mahasiswa
     UNION ALL
     SELECT `NIM` AS jenis_data, nim AS data, mhs_id
     FROM mahasiswa
     UNION ALL
     SELECT `Nama Lengkap` AS jenis_data, nama AS data, mhs_id
     FROM mahasiswa
) AS data_akademik_inner
GROUP BY jenis_data
) AS data_akademik_outer
GROUP BY jenis_data;
```

Figure 6. Airflow program code snippet to pull data from SALAM application database to datamart on completeness dimension

The above program code snippet is a query on the completeness dimension. In general, the query pulls data from the Salam application according to its data quality measurement attributes as data types. Each type of data will be calculated for the total data, the total data filled and unfilled. The data is then calculated for its DPMO which can then measure the quality of academic data from measurements using the Six Sigma method.

source code airflow untuk dimensi uniqueness

```sql
SELECT jenis_data,
SUM(total_data) AS total_data,
SUM(total_data_ganda) AS total_data_ganda,
DPMO,
CASE
    WHEN DPMO <= 3.4 THEN 6
    WHEN DPMO <= 5 THEN 5.9
    WHEN DPMO <= 9 THEN 5.8
    WHEN DPMO <= 13 THEN 5.7
    WHEN DPMO <= 21 THEN 5.6
    --aturan DPMO lainnya--
    ELSE 0
END AS sigma_metrics
FROM (
SELECT
 Jenis_Data,
 SUM(Total_Data) AS Total_Data,
 SUM(Total_Data_Ganda) AS Total_Data_Ganda,
 SUM(DPMO) AS DPMO
FROM (
 SELECT
    'ID Mahasiswa' AS Jenis_Data,
    COUNT(*) AS Total_Data,
    (
        SELECT COUNT(*)
        FROM (
            SELECT COUNT(*) AS count
            FROM mahasiswa
            GROUP BY mhs_id
            HAVING count > 1
        ) AS duplicates
    ) AS Total_Data_Ganda,
    (CAST((
        SELECT COUNT(*)
        FROM (
            SELECT COUNT(*) AS count
            FROM mahasiswa
            GROUP BY mhs_id
            HAVING count > 1
        ) AS duplicates
    ) AS DECIMAL) / COUNT(*) * 1000000) AS DPMO
 FROM mahasiswa
) AS subquery
UNION ALL
SELECT
 'NIM' AS Jenis_Data,
 COUNT(*) AS Total_Data,
 (
    SELECT COUNT(*)
    FROM (
        SELECT COUNT(*) AS count
        FROM mahasiswa
        GROUP BY nim
        HAVING count > 1
    ) AS duplicates
 ) AS Total_Data_Ganda,
 (CAST((
    SELECT COUNT(*)
    FROM (
        SELECT COUNT(*) AS count
        FROM mahasiswa
        GROUP BY nim
        HAVING count > 1
    ) AS duplicates
 ) AS DECIMAL) / COUNT(*) * 1000000) AS DPMO
FROM mahasiswa
) AS main_query
GROUP BY jenis_data;
```

Figure 7. Airflow program code snippet to pull data from SALAM application database to datamart on uniqueness dimension

The program code snippet in the image above is the airflow program code that functions to pull unique data from the SALAM application. Academic data that has a certain level of uniqueness such as student ID and NIM should not have duplicate data.

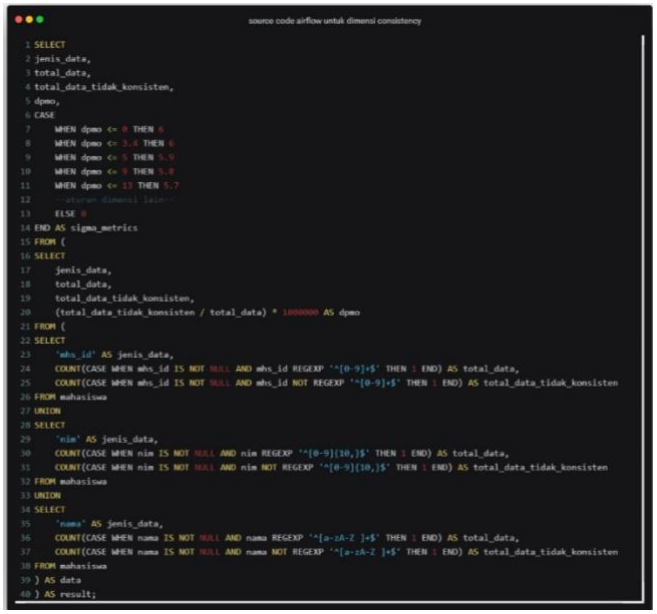

Figure 8. Airflow program code snippet to pull data from SALAM application database to datamart on consistency dimension

Each attribute is also measured for data quality related to its consistency. Each data attribute is checked to see whether the data is in accordance with its criteria or not. For example, NIM data must consist of numbers between 0 and 9 and at least consist of 10 digits as in the previous image.

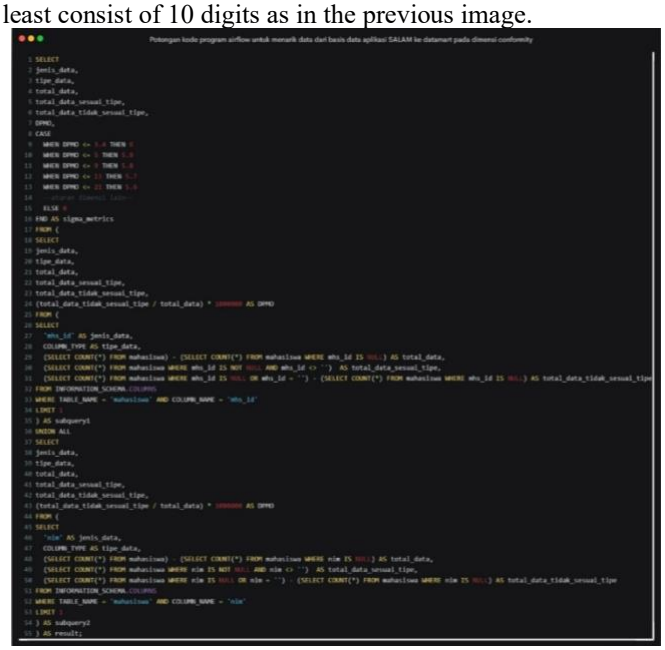

Figure 9. Airflow program code snippet to pull data from SALAM application database to datamart in conformity dimension

The image above is a piece of code to measure data quality based on conformity with the parameter of data defects that do not match the type of data filled in an attribute. Measurement is done by comparing the total data filled in that matches the data type, and the total data filled in that does not match the data type to get the DPMO value that determines the sigma metric value of the attribute.

Potongan kode program airflow untuk menarik data dari basis data aplikasi SALAM ke datamart pada dimensi timeliness

```
SELECT
mulai_smt,
CASE WHEN COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) = 0 THEN 'Tersedia' ELSE '' END AS status,
(COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) / COUNT(mhs_id)) * 1000000 AS DPMO,
CASE
  WHEN (COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) / COUNT(mhs_id)) * 1000000 <= 3.4 THEN 6
  WHEN (COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) / COUNT(mhs_id)) * 1000000 <= 5 THEN 5.9
  WHEN (COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) / COUNT(mhs_id)) * 1000000 <= 9 THEN 5.8
  WHEN (COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) / COUNT(mhs_id)) * 1000000 <= 13 THEN 5.7
  WHEN (COUNT(CASE WHEN mhs_id IS NULL THEN 1 ELSE NULL END) / COUNT(mhs_id)) * 1000000 <= 21 THEN 5.6
--aturan lainnya--
  ELSE 0
END AS sigma_metrics
FROM
mahasiswa
GROUP BY
mulai_smt
ORDER BY
mulai_smt;
```

Figure 10. Airflow program code snippet to pull data from SALAM application database to datamart on timeliness dimension

The last measurement standard is based on the timeliness dimension. Measurements are made to measure the availability of academic data on the SALAM application whether it is available from the last 7 years with the parameter of the availability of student graduation limits of 14 semesters (7 years). Each program code from each dimension will then be run using the DAG (Directed Acyclic Graph) program code using the Python programming language. The DAG program code is run periodically every 24 hours so that the control process of the quality of academic data on the SALAM application can be seen periodically.

```
from airflow import DAG
from datetime import timedelta
from datetime import datetime
from airflow.operators.bash import BashOperator
from airflow.utils.dates import days_ago
default_args = {
    'depends_on_past': False,
}
with DAG(
    'dq_completeness_biodata_mahasiswa',
    default_args=default_args,
    description='Kualitas Data Biodata Mahasiswa berdasarkan kelengkapannya',
    schedule_interval='@daily',
    start_date=datetime(2023, 4, 26),
    tags=['SALAM', 'Data_Quality', 'Biodata_Mahasiswa', 'completeness'],
    catchup=False,
) as dag:
    t1 = BashOperator(
        task_id='get_data',
        bash_command='php /root/php-scripts/salam/dq_completeness_biodata_mahasiswa.php',
        dag=dag,
    )
    t1
```

Figure 11. DAG (Directed Acyclic Graph) program code structure to run the program code structure for retrieving data from the SALAM application using apache airflow to the datamart database

The program code imports several modules and classes from the Apache airflow library to run a DAG (Directed Acyclic Graph) workflow consisting of several tasks. The first thing to do is to import modules and classes, namely DAG from the airflow module which functions to create DAG objects that represent workflows, timedelta and datetime modules from the datetime library. This code is used to set the time and interval in the DAG, import the BashOperator class from the bash module to be used to run bash commands in the DAG task, and import the days_ago function from the dates module to calculate the date based on the past day. Next, the default arguments configuration is carried out for variables that contain the default configuration for the DAG. In this example, there is only one configuration, namely 'depends_on_past': False, which indicates that the task does not depend on the previous task. Next, the DAG is created with the with DAG(...) as DAG:: function which defines the DAG object according to the required DAG.

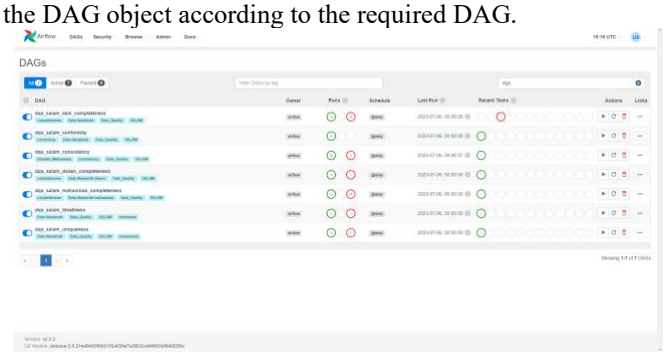

Figure 12. List of DAGs on academic data quality measurement on apache airflow

Each dimension in the academic data quality measurement process has different DAGs. There are 7 DAGs for 5 dimensions used in the academic data quality measurement process in the SALAM application. Each DAG will run the airflow program code from each dimension in the data quality measurement.

3.2 Analyze

Analyze is the next step after the define and measure process. The analysis process is a critical stage to identify the root cause of the problem and evaluate the measurement data according to the dimensions used. This process involves collecting data on the process being analyzed, DPMO results to sigma metrics from the results of measuring the quality of academic data in the SALAM application. After data retrieval using apache airflow, the results of measuring the quality of academic data can be seen in the destination table according to what is defined in the airflow program code in each data quality measurement dimension. Figure 4.10 shows an example of the display in the Navicat application for the results of measuring the quality of student academic data for the completeness dimension. Each dimension of measuring the quality of academic data will appear in each table according to the DAG created.

Table 2. Analysis of the average results of data quality measurements on academic data in the SALAM application using the Six Sigma method

No.	Dimension	DPMO	Sigma Metric
1.	*Completeness*	214826.66	3.57
2.	*Uniqueness*	245234.29	3.64
3.	*Consistency*	203207.52	3.82
4.	*Conformity*	4807.96	4.89
5.	*Timeliness*	0.00	6.00

The results of measuring the quality of academic data in the SALAM application against 129 data attributes show different results in each measurement dimension. The results of these measurements can be seen in table 4.3 regarding the analysis of the average results of data quality measurements on academic data in the SALAM application.

The consistency dimension has a DPMO score of 214826.66 with a sigma metric score of 3.57. Measurements with this dimension indicate that the academic data in the Salam application is quite good when viewed from the consistency of the data.

The completeness dimension has a DPMO score of 245234.29 with a sigma metric score of 3.64, which means that the dimension occupies sigma level 3 with a fairly low level of defects.

The results of measurements on the conformity dimension show sigma level 3 results with a DPMO value of 203207.52 with a sigma metric value of 3.82. In this dimension, the quality of academic data in the Salam application has a significant increase in quality but still allows for errors that are acceptable in a certain amount.

For the uniqueness dimension, the results of measuring the quality of academic data in the SALAM application show quite good results with a sigma level 4 category. This dimension has a DPMO score of 4807.96 with a sigma metric score of 4.89 with a low level of defects.

The timeliness dimension is a measurement parameter that has a perfect sigma metric score. This dimension has a DPMO score of 0 with a sigma metric score of 6, which means that when measured from this dimension, academic data in the SALAM application has a high level of quality, almost no defect level, and a very high level of excellence.

Overall, the consistency dimension greatly influences the poor quality of academic data in the SALAM application with a sigma metric score of 3.57. Meanwhile, the good quality of academic data in the SALAM application is influenced by the availability of data from year to year. The timeliness dimension has a perfect sigma metric score, which is 6. The quality of academic data in the Salam application shows that there is still a possibility of errors that can be accepted in quite large numbers, but on the other hand, the quality of academic data in the Salam application has increased quite significantly.

3.3 Improvement

After the process of defining, measuring and analyzing, the next step is the improve process which focuses on the development and implementation of solutions designed to address the problems identified during the previous stage. At this stage, several improvement solutions will be designed that can be proposed to improve the quality of academic data in the SALAM application.

Figure 13. Average quality of academic data on the SALAM application based on sigma metric levels

The image above shows the results of measuring the quality of academic data on the SALAM application based on the sigma metric obtained. From these results, it can be seen that consistency ranks first when viewed from the average value of its DPMO. This is followed by the dimensions of completeness, conformity, uniqueness and timeliness. Each dimension has its own DPMO value with the cause of the attributes in it that are measured so that they have different sigma metrics.

3.4 Control

After the improvement process is carried out, which focuses on the development and implementation of improvement solutions that can be proposed to improve the quality of academic data in the SALAM application. The last step in the DMAIC method is the control process which aims to ensure that the improvements that have been made are maintained and provide sustainable results. This process involves the development and implementation of control steps that will ensure that the improved process continues to run consistently and produces the expected results. The following are the control focuses that can be carried out.

1. Develop Control Plan (Develop Control Plan). In order to be able to control the value of the quality of academic data owned by UIN Sunan Gunung Djati Bandung to be better, it is necessary to develop a control plan to control the value of data quality to be even better.
2. Establish Monitoring Systems (Build Monitoring Systems). The monitoring system has been conceptualized in the visualization of the improvement process data. The monitoring system can later be integrated with the executive dashboard of the leadership (TERAS UIN SGD BANDUNG).
3. Implement Process Controls (Implementing Process Controls). The process of improving academic data quality control must be routinely monitored and provide periodic warnings to the academic community to update and complete the data contained in the SALAM application.

Continuous Monitoring and Improvement (Continuous Monitoring and Improvement). Continuous monitoring and improvement can be carried out periodically every day because the academic data whose data quality is measured has been arranged to be pulled periodically every day using apache airflow. In addition to monitoring, there must be continuous improvements made in the form of periodic data updates so that the data owned will be increasingly up to date.

4. CONCLUSION

This study uses the Six Sigma method with the DMAIC methodology to give an objective and quantifiable assessment of academic data quality in the SALAM application, based on the goals stated in the introduction. Completeness, uniqueness, consistency, conformance, and timeliness are the five main characteristics of data quality that are evaluated. These dimensions are measured using DPMO and sigma metrics to represent the true state of the data. The findings emphasize both the advantages and disadvantages of the existing academic data management procedures, show different degrees of data quality across dimensions, and pinpoint the main causes of flaws. These results validate Six Sigma's suitability as a methodical framework for evaluating and enhancing the quality of academic data in higher education establishments and provide a basis for suggesting focused improvement and control measures. However, the results may not be as broadly applicable as they may be because this study is restricted to a specific institutional application and depends on predetermined data quality dimensions. In order to support more proactive and data-driven decision making in higher education environments, future research is encouraged to broaden the scope to include other institutional systems, create automated and continuous data quality monitoring mechanisms, incorporate quality indicators into executive dashboards, and investigate advanced analytics or machine learning techniques.

ACKNOWLEDGEMENTS

After measuring the quality of academic data on the SALAM application using the Six sigma method, the following conclusions can be drawn.

1. The level of academic data quality on the SALAM application has a fairly good level of data quality with a sigma metric score of 4.38.
2. Based on the 5 dimensions used as standards for measuring data quality, the consistency and completeness dimensions are dimensions that greatly affect the poor quality of academic data on the SALAM application so that data updates are needed so that the quality of academic data owned by UIN Sunan Gunung Djati Bandung is getting better.

REFERENCES

[1] I. Anas and A. Hidayat, "Implementasi Algoritma Vigeneere Cipher dan GOST dalam Keamanan Data," *Journal & Penelitian Teknik Informatika*, vol. 2, no. 2, 2018.
[2] Jr. , C. G. C. R. Kelly Rainer, *Introduction to information systems*. 2013.
[3] D. Sawitri, "REVOLUSI INDUSTRI 4.0 : BIG DATA MENJAWAB TANTANGAN REVOLUSI INDUSTRI 4.0," 2019.
[4] B. , H. D. , K. M. , S. A. , & S. M. Heinrich, "Requirements for Data Quality Metrics," *Journal of Data and Information Quality*, vol. 9, no. 2, pp. 1–32, 2018.
[5] "Peluncuran Aplikasi New SALAM pada Rapat Koordinasi Bagian Akademik Tahun 2021 - UIN Sunan Gunung Djati Bandung." Accessed: Jun. 16, 2023. [Online]. Available: https://uinsgd.ac.id/peluncuran-aplikasi-new-salam-pada-rapat-koordinasi-bagian-akademik-tahun-2021/
[6] N. Makhoul, "Review of data quality indicators and metrics, and suggestions for indicators and metrics for structural health monitoring," Dec. 01, 2022, *Springer*. doi: 10.1186/s43251-022-00068-9.
[7] D. Gürdür, J. El-khoury, and M. Nyberg, "Methodology for linked enterprise data quality assessment through information visualizations," *J Ind Inf Integr*, vol. 15, pp. 191–200, Sep. 2019, doi: 10.1016/j.jii.2018.11.002.
[8] "Analisis Penerapan ISO 9001:2015 Industri Makanan: Sebuah Narrative Literature Review", [Online]. Available: https://jisma.org
[9] "ANALISIS MANAJEMEN MUTU TERPADU (TQM) DALAM PELAYANAN RUMAH SAKIT."
[10] B. Klefsjo, H. Ê. Kan Wiklund, and R. L. Edgeman, "SIX SIGMA SEEN AS A METHODOLOGY FOR TOTAL QUALITY MANAGEMENT." [Online]. Available: http://www.emerald-library.com/ft
[11] S. Koppel and S. Chang, "MDAIC – a Six Sigma implementation strategy in big data environments," *International Journal of Lean Six Sigma*, vol. 12, no. 2, pp. 432–449, Mar. 2021, doi: 10.1108/IJLSS-12-2019-0123.
[12] A. S. Patel and K. M. Patel, "Critical review of literature on Lean Six Sigma methodology," 2020, *Emerald Group Holdings Ltd.* doi: 10.1108/IJLSS-04-2020-0043.
[13] Y. Z. Mehrjerdi, "Six-Sigma: Methodology, tools and its future," *Assembly Automation*, vol. 31, no. 1, pp. 79–88, 2011, doi: 10.1108/01445151111104209.
[14] Y. Latief and dan Retyaning Puji Utami, "PENERAPAN PENDEKATAN METODE SIX SIGMA DALAM PENJAGAAN KUALITAS PADA PROYEK KONSTRUKSI," 2009.
[15] S. H. Cahyono, Y. G. Sucahyo, J. S. Raya, and J. Pusat, "Pengukuran Kualitas Data Menggunakan Framework Total Data Quality Management (TDQM): Studi Kasus Sistem Informasi Beasiswa Universitas Indonesia Data Quality Assessment Using the TDQM Framework: A Case Study of University of Indonesia (UI) Scholarship Information System," *Jurnal Ilmu Pengetahuan dan Teknologi Komunikasi*, vol. 22, no. 2, pp. 193–206, doi: 10.33164/iptekkom.22.2.2020.193-206.
[16] D. Nur Fauziah, D. Ayu Nur Wulandari, S. Informasi, K. Akuntansi, S. Nusa Mandiri Jakarta, and A. BSI Karawang, "PENGUKURAN KUALITAS LAYANAN BUKALAPAK.COM TERHADAP KEPUASAN KONSUMEN DENGAN METODE WEBQUAL 4.0," 2021, [Online]. Available: http://www.nusamandiri.ac.id1,http://www.bsi.ac.id2
[17] M. Panwar, "Application of Six-Sigma for Data Quality Improvement in an Insurance Company," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 4, no. 09, 2015, [Online]. Available: www.ijstr.org
[18] J. De Mast and J. Lokkerbol, "An analysis of the Six Sigma DMAIC method from the perspective of problem solving," *Int J Prod Econ*, vol. 139, no. 2, pp. 604–614, Oct. 2012, doi: 10.1016/j.ijpe.2012.05.035.
[19] M. J. Jamshidi, M. Hosseinpour, H. Heshmati, and B. Fathi Zolmabadi, "Improving the Performance of Hospital Information Systems Using Six Sigma for Kermanshah Province Hospitals," *Journal of Clinical Research in Paramedical Sciences*, vol. 10, no. 1, Mar. 2021, doi: 10.5812/jcrps.102448.
[20] P. Kaushik and D. Khanduja, "Application of six sigma DMAIC methodology in thermal power plants: A case study," *Total Quality Management and Business Excellence*, vol. 20, no. 2, pp. 197–207, 2009, doi: 10.1080/14783360802622995.
[21] A. M. Ponsiglione *et al.*, "A six sigma DMAIC methodology as a support tool for health technology assessment of two antibiotics," *Mathematical Biosciences and Engineering*, vol. 18, no. 4, pp. 3469–3490, 2021, doi: 10.3934/MBE.2021174.
[22] M. M. B. Tufail, A. Shamim, A. Ali, M. Ibrahim, D. Mehdi, and W. Nawaz, "DMAIC methodology for achieving public satisfaction with health departments in various districts of Punjab and optimizing CT scan patient load in urban city hospitals," *AIMS Public Health*, vol. 9, no. 2, pp. 440–457, 2022, doi: 10.3934/publichealth.2022030