Model business intelligence system design of quality products by using data mining in R Bakery Company

R Fitriana¹, J Saragih and N Luthfiana

System & Simulation Industry Laboratory, Industrial Engineering Department, Trisakti University, Jakarta, Indonesia
¹rinaf@trisakti.ac.id

Abstract. R Bakery company is a company that produces bread every day. Products that produced in that company have many different types of bread. Products are made in the form of sweet bread and wheat bread which have different tastes for every types of bread. During the making process, there were defects in the products which the defective product turns into reject product. Types of defects that are produced include burnt, sodden bread and shapeless bread. To find out the information about the defects that have been produced then by applying a designed model business intelligence system to create database and data warehouse. By using model business Intelligence system, it will generate useful information such as how many defect that produced by each of the bakery products. To make it easier to obtain such information, it can be done by using data mining method which data that we get is deep explored. The method of data mining is using k-means clustering method. The results of this intelligence business model system are cluster 1 with little amount of defect, cluster 2 with medium amount of defect and cluster 3 with high amount of defect. From OLAP Cube method can be seen that the defect generated during the 7 months period of 96,744 pieces.

Keywords: Model, Business Intelligence System, Data Mining, K-Means Clustering

1. Introduction

Nowadays competition becomes more intense because competitors in the same field has been increased. Many of companies flocked to improve their quality product in order to attract customers to buy their products, so did R Bakery. To face the competition with other bread company, R Bakery should be able to pay attention of the weakness factors so the company can overcome these weaknesses and be superior compared with the other competitors. The factors that R Bakery need to consider is the quality of products and how many defects of product that have been produced.

The application of model business intelligence system can make benefit for the company because business intelligence is a technique that used to transform raw data into useful data in the future. Other than that business intelligence can identify unstructured data, can improve the quality of products and create new opportunities business strategies. The purposes of this research are designing a model business intelligence system in R Bakery company and designing data mining models by using k-means clustering method.

Business Intelligence is the extraction process of operational data and collected in a datawarehouse [2]. There are many researches about business intelligence system. Business Intelligence Model in customer relationship management and quality for dairy agroindustry was made as integration of Cube, Data warehouse Model and Fuzzy system and it helps for a fast and efficient transaction in the system [7]. The topic that integrated with BI research is Supply Chain Management, Customer Relationship Management, Data Mining, Data Warehouse, Decision Support System, Performance
Scorecard, Knowledge Management, Business Process Management, Artificial Intelligence, Enterprise Resource Planning, Extract Transformation Loading, OLAP, Quality Management System, Strategic Management. The role of business intelligence systems was to enhance customer satisfaction and to develop decision support tools to improve customer satisfaction by the service provided [8]. A research consists of object-oriented analysis and design for data warehouse development in the business intelligence system has made. Modelling the core business had supported by financing and quality management model to support a modelling approach to computing intelligence and fuzzy system. [10]. State of the art of this research is the integration between business intelligence, data mining, olap cube in bread company.

The difference between the quality of data and information quality is very evident in business intelligence system. The main goal of business intelligence is to provide the quality of information for decision making [12]. Clustering is a collection of the records that have a similarity with each other of data and have dissimilarity with the record in another cluster. To do this clustering is by using k-means algorithm [5]. K-means is one of the simplest methods of grouping. The procedure is simple easy to classify the data provided through a number of clusters [4]. Distance calculation is using the Euclidean formula as follows:

\[ d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  

Where: \( n \) = number of data; \( x_i = \text{data}-i \); \( y_i = \text{centroid}-i \)

Unified Modelling Language (UML) is a standard language to create a software design [1]. ETL (Extract, Transform, Load) process is a process that needed to make the data that ready for used in data warehouse [6]. Data warehouses are the collections of data that have the character of the subject-oriented, integrated, unchanged and have a particular time frame that supports management decision [3].

2. Research Method

![Figure 1. Methodology Flowchart](image-url)
3. Result and Analysis

3.1. System Analysis
To analyse the need of system can be made by using PIECES method. PIECES method is used to analyse requirements and solutions as well [11]. Aspects that analysed are performance, information, economic, control, efficiency and customer service. This is table PIECES analysis bread company R Bakery.

| PIECES   | Old System                                                                 | New System                                                                 |
|----------|----------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Performance | Time needed to analyze the number of defects generated long enough because the recording is still done manually | Data processing in the form of display the number of the defects product produced faster because it is processed on based computerized |
| Information | The result of information is less accurate and relevant because of frequent errors and mistake during the process of determining the amount defects are generated | More accurate and relevant because did not occur because many of mistake error in the processing data |
| Economic   | Recording is still done manually                                           | To spend money to buy computer and training Employee to used software      |
| Control    | Control safety is less because data that produces still shaped file        | Easy to control, safety control                                           |
|            | There is potential human error.                                            | Can help in making decision making With fast and right                    |
| Efficiency | Need extra cost and time.                                                 | More economical, effective and efficient                                  |
| Service    | Service and searching data less fast because have to open file.           | Service and collect data can be faster because using computerized system   |

System analysis in this research is doing the interview and collecting production data of R Bakery company. From the interviews, it was found that the bakery has not had a database and information system to monitor the amount defect of bread when the production takes place so researches set out to create a database and information systems with design an intelligence business system about the quality of products in R Bakery. The data that retrieved is defective products for 7 months. Here’s an overview of business intelligence system design.

![Figure 2. Business Intelligence System](image-url)
3.2. **UML (Unified Modelling Language)**

UML is a stage design that aims to make easier to understand so when designing the system people know of how the system will be created.

![Figure 3. Clustering Activity Diagram](image)

In the data mining activity diagram above, the employee in charge of production in gathering the required data for the establishment of intelligence business system. The required data are the amount of bread and type of defects that had been produced during 7 (months) period starting from June until December of 2016. The data that had been collected then given to the production head that in charge of making intelligence business systems. The method that used to make the system is data mining clustering by using Weka Software.

3.3. **Making Data warehouse and Model Business Intelligence System**

While making draft of intelligence business system, the first step that must be done is to design a data warehouse that will be created. To create the design of data warehouse is by create a snowflake schema as shown below:

![Figure 4. R Bakery Snowflake Schema](image)
From the picture above that in making data warehouse consists of dimensions of time, product, defect, master categories and master of defects which merged later into the data warehouse production which containing a combination of attributes of each dimension. Here’s the overview of product dimension and defect dimension that executed by using software called spoon.

![Spoon Product Dimension](image)

**Figure 5. Spoon Product Dimension**

In designing intelligence business system, the required database is product dimension, time dimension, defect dimension and fact defect. Fact defect is a merger between the product dimension, time dimension and defect dimension. As seen in the image above, the picture is the visualization of fact defect that contains data information about production data and the number of defects that had been produced during the production process.

3.4. Data Mining

Data mining aims to obtain information from various data sources to make easier in decision making. Data mining method is using k-means clustering method. The use of k-means clustering is that by classifying the product defects during the production process from June until December of 2016 consisting of three clusters include a cluster 1 with little amount of defect, cluster 2 with medium amount of defect and cluster 3 with high amount of defect.

| Table 2. R Bakery’s Product Defect |
|-----------------------------------|
| **Product Code** | **Product Name** | **Burnt** | **Sodden** | **Shapeless** | **Product Code** | **Product Name** | **Burnt** | **Sodden** | **Shapeless** |
| PD-1 | Chocolate | 2998 | 2380 | 2591 | PD-13 | Beef | 933 | 941 | 947 |
| PD-2 | Coconut | 1462 | 1230 | 1211 | PD-14 | Sausage | 800 | 754 | 667 |
| PD-3 | Sardine | 1319 | 1303 | 1244 | PD-15 | Milk | 1192 | 1072 | 1068 |
| PD-4 | Mocha | 1147 | 1082 | 1001 | PD-16 | 5 Flavors | 604 | 524 | 549 |
| PD-5 | Donuts | 3095 | 2656 | 2554 | PD-17 | Strawberry | 721 | 649 | 726 |
| PD-6 | Jelly | 657 | 567 | 518 | PD-18 | Sausage Shredded | 449 | 421 | 401 |
| PD-7 | Satay | 1382 | 1239 | 1344 | PD-19 | KPK | 566 | 506 | 512 |
| PD-8 | Choco Banana | 1051 | 907 | 894 | PD-20 | Square Set | 1542 | 1341 | 1474 |
| PD-9 | Choco Cheese | 1184 | 1103 | 1190 | PD-21 | Small Set | 1200 | 1149 | 1061 |
| PD-10 | Cheese Banana | 2306 | 2074 | 2017 | PD-22 | Pee Bread | 2819 | 2888 | 2931 |
| PD-11 | Blueberry | 117 | 69 | 35 | PD-23 | Wheat | 5012 | 4085 | 4109 |

Example of Iteration 1:
Determine First Centre of Cluster

Assumption of first centre:
- Taken data-11 as centre of cluster 1: (117, 69, 35)
- Taken data-5 as centre of cluster 2: (3095, 2656, 2554)
- Taken data-23 as centre of cluster 3: (5012, 4085, 4109)

Calculation Centroid of cluster

to calculate the distance between data and centroid by using Euclidean formula (1) as follows:

\[ d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]

\[ PD-1 = \sqrt{(2998-117)^2 + (2380-69)^2 + (2591-35)^2} = 4991.549621 \]

\[ PD-2 = \sqrt{(1462-117)^2 + (1249-69)^2 + (1211-35)^2} = 2141.121435 \]

The number of iterations as many as three iterations. The calculation to determine the centroid of cluster was the same as before, the difference of distance of the cluster is the average value for each cluster, so the result as follows:

| Product Code | Product Name          | Burnt  | Sodden | Shapeless | Cluster |
|--------------|-----------------------|--------|--------|-----------|---------|
| PD-1         | Chocolate             | 2998   | 2380   | 2591      | Cluster 2 |
| PD-2         | Coconut               | 1462   | 1249   | 1211      | Cluster 1 |
| PD-3         | Sarikaya              | 1319   | 1303   | 1244      | Cluster 1 |
| PD-4         | Mocha                 | 1147   | 1082   | 1001      | Cluster 1 |
| PD-5         | Donuts                | 3095   | 2656   | 2554      | Cluster 2 |
| PD-6         | Jelly                 | 657    | 567    | 518       | Cluster 1 |
| PD-7         | Satay                 | 1382   | 1239   | 1344      | Cluster 1 |
| PD-8         | Choco Banana          | 1051   | 907    | 894       | Cluster 1 |
| PD-9         | Cheese                | 1184   | 1103   | 1190      | Cluster 1 |
| PD-10        | Choco Banana Cheese   | 2306   | 2074   | 2017      | Cluster 2 |
| PD-11        | Blueberry             | 117    | 69     | 35        | Cluster 1 |
| PD-12        | Choco Cheese          | 1925   | 1872   | 1861      | Cluster 2 |
| PD-13        | Beef                  | 933    | 941    | 947       | Cluster 1 |
| PD-14        | Sausage               | 800    | 754    | 667       | Cluster 1 |
| PD-15        | Milk                  | 1192   | 1072   | 1068      | Cluster 1 |
| PD-16        | 5 Flavors             | 604    | 524    | 549       | Cluster 1 |
| PD-17        | Strawberry            | 721    | 649    | 726       | Cluster 1 |
| PD-18        | Sausage Shredded      | 449    | 421    | 401       | Cluster 1 |
| PD-19        | KPK                   | 566    | 506    | 512       | Cluster 1 |
| PD-20        | Square Set            | 1542   | 1341   | 1474      | Cluster 1 |
| PD-21        | Small Set             | 1200   | 1149   | 1061      | Cluster 1 |
| PD-22        | Peel Bread            | 2819   | 2888   | 2931      | Cluster 2 |
| PD-23        | Wheat                 | 5012   | 4085   | 4109      | Cluster 3 |

3.5. Processing OLAP Cube

The database used are defect dimensions, product dimensions and time dimensions. Schema depiction can be seen in the following figure.
From the schema is then applied to software called Pentaho Analysis. Here’s the look of Pentaho Analysis based on the schema workbench that has been created.

In Pentaho Analysis image above can be seen that the defect generated during the 7 months period of 96,744 pieces. Broken bread products must be given special attention because if it was left unchecked the greater the losses suffered by the company. The occurrence of the number of defects in the product can be caused by several factors:

1. Skill on human resources in the production of less good
2. Supervision on part of the production is not good.

4. Conclusion

1. The model business intelligence system is made with the Extract Transform Load process with model of data mining to improve quality product in bread factory.
2. The method of data mining is using k-means clustering method. The results of this intelligence business model system are cluster 1 with little amount of defect, cluster 2 with medium
amount of defect and cluster 3 with high amount of defect. Cluster 0 is a product with little amount of defect, cluster 1 with medium amount of defect and cluster 2 with high amount of defect. Products that are in the cluster 0 include jelly, blueberry, sausage, 5 flavors, strawberry, sausage shredded and KPK. Products that are in the cluster 1 include coconut bread, sarikaya, mocha, satay, choco banana, cheese, choco cheese, beef, milk, square set and small set. Products that are in cluster 2 include chocolate, donuts, choco banana cheese, peel bread and wheat bread.

3. From OLAP Cube Method can be seen that the defect generated during the 7 months period of 96,744 pieces.

5. References
[1] Booch, James G, et.al 2005 The unified modeling language user guide 2nd edition (United State: Addison Wesley Proffesional).
[2] Imelda 2011. Business intelligence. Majalah Ilmiah UNIKOM 2011, Vol. 11, No. 1.
[3] Inmon, W H 2002 Building the data warehouse. 3rd edition. USA: John Willey & Sons Inc.
[4] Kristanto, Christopher H, et.al 2016 Implementasi k-means clustering untuk pengelompokan analisis rasio profitabilitas dalam working capital. JUISI, Vol. 02, No.01.
[5] Larose D 2006 Data mining methods and models. (New Jersey: John Willey & Sons Inc).
[6] Rainardi V 2007 Building a data warehouse with examples in sql server. APRESS.
[7] Fitriana R, Eriyatno, Djinna T, and Kusmuljono B S 2012 Peran sistem intelijensia bisnis dalam manajemen pengelolaan pelanggan dan mutu untuk agroindustri susu skala usaha menengah. J. Teknologi Industri Pertanian 22 (3): 131 – 139.
[8] Fitriana R, Saragih J and Andhika M 2016 Business intelligence system models proposals to improve the quality of service at pt. gia. Proc. 9th Int. Seminar Industrial Engineering and Management.
[9] Fitriana R, Eriyatno and Djinna T 2011 Progress in business intelligence research: a literature review. Int. J. of Basic & Applied Sciences IJBAS-IJENS 11(3).
[10] Fitriana R, and Djinna T 2012 Business intelligence design for decision support dairy agro industry medium scaled enterprise. Int. J. of Engineering & Technology IJET-IJENS 12(5)
[11] Whitten, Bentley, Dittmand. 2004. System analysis and design methods. 6th edition. The McGraw-Hill Companies.
[12] Wieder B and Luise M. 2015 The impact of business intelligence on the quality of decision making – a mediation model. (Australia: Elsevier B.V.)