Bulletin of Electrical Engineering and Informatics
Vol. 12, No. 5, October 2023, pp. 3016~3026
ISSN: 2302-9285, DOI: 10.11591/eei.v12i5.5162

Predictive business intelligence dashboard for food and beverage business

Nathaniel Wikamulia, Sani Muhamad Isa
Department of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Nov 7, 2022
Revised Dec 16, 2022
Accepted Jan 29, 2023

Keywords:

Business intelligence dashboard
Customer segmentation
Demand forecasting
Predictive model
Recency, frequency, and monetary analysis

ABSTRACT

This research was conducted to provide an example of predictive business intelligence (BI) dashboard implementation for the food and beverage business (businesses that sell fast-expired goods). This research was conducted using data from a bakery's transactional database. The data are used to perform demand forecasting using extreme gradient boosting (XGBoost), and recency, frequency, and monetary value (RFM) analysis using mini batch k-means (MBKM). The data are processed and displayed in a BI dashboard created using Microsoft Power BI. The XGBoost model created resulted in a root mean square error (RMSE) value of 0.188 and an R2 score of 0.931. The MBKM model created resulted in a Dunn index value of 0.4264, a silhouette score value of 0.4421, and a Davies-Bouldin index value of 0.8327. After the BI dashboard is evaluated by the end user using a questionnaire, the BI dashboard gets a final score of 4.77 out of 5. From the BI dashboard evaluation, it was concluded that the predictive BI dashboard succeeded in helping the analysis process in the bakery business by: accelerating the decision-making process, implementing a data-driven decision-making system, and helping businesses discover new insights.

Corresponding Author:

Nathaniel Wikamulia
Department of Computer Science, Bina Nusantara University
JL. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
Email: nathaniel.wikamulia@binus.ac.id

1. INTRODUCTION

Business intelligence dashboard (BI dashboard) is a well-known data visualization method due to its ability in discovering insights and speeding up decision-making process [1]. The advantages of BI dashboard are evident from research conducted by Hansoti [2] where it was found that departments within a company can make decisions more quickly and more data-driven with BI dashboard. In addition, BI dashboard is also widely used because it can perform data visualization for various cases such as banking, healthcare, education, food and beverage, and other cases.

While BI dashboard can help various businesses, the BI dashboard analysis created for one business is not applicable for other businesses. For example, the food and beverage business can only make short term predictions because its business trends change quickly. Due to the differences in the characteristics of the food and beverage business with other businesses, this study aims to create a predictive BI dashboard (BI dashboard that can make predictions) that can provide analysis for the food and beverage business and other businesses whose products are perishable.

There are several similar studies to this research. Diana *et al.* [3] created a predictive BI dashboard to predict the final grades of students and to monitor the learning conditions of the students. Hassanudin *et al.* [4] made a predictive BI dashboard to monitor corrosion rate on pipelines using artificial neural network (ANN) and multiple linear regression analysis (MLRA). Cho *et al*. [5] created a predictive BI dashboard to predict factory machine's condition. Dagliati *et al*. [6] created a predictive BI dashboard to predict a patient's condition and provide recommendations for treatment that must be done to the patient.

Although these studies have succeeded in creating a predictive BI dashboard, there is no predictive BI dashboard to analyze the food and beverage business (businesses that sell fast-expired goods). Therefore, the main contribution of this research is to create a predictive BI dashboard that can help the decision-making process for the food and beverage business. This research will provide an example of predictive BI dashboard implementation for the food and beverage business using a dataset from a bakery. The machine learning algorithms that are used in this research are extreme gradient boosting (XGBoost) and mini batch k-means (MBKM). XGBoost is used to perform demand forecasting on the amount of bread that the bakery must produce. MBKM algorithm is used to categorize customers based on the recency, frequency, and monetary (RFM) value of each customer.

The remainder of this paper is organized as follows. Section 2 describes the state-of-the-art methods for implementing the predictive BI dashboards, demand forecasting model, and RFM analysis model. Section 3 provides an overview of the predictive BI dashboards, XGBoost model, and MBKM model development. Section 4 describes the research method. Section 5 shows the evaluation results and analysis of the predictive BI dashboard. Section 6 shows the conclusion of the research.

2. RELATED WORKS

This section is divided into three sub sections, subsection 2.1 explain several research about predictive BI dashboard implementation. Subsection 2.2 explain several research on demand forecasting using machine learning algorithms. Subsection 2.3 explain several research on methods to perform RFM analysis using machine learning.

2.1. Predictive business intelligence dashboard

Predictive BI dashboard is a topic that is often researched by various organizations due to its ability to predict future conditions and patterns [7]. There are several research about predictive BI dashboard implementation. Diana *et al*. [3] created a predictive BI dashboard that displays information on student's learning status and predict the student's final grade from log data. The prediction is performed using a supervised machine learning model. With this BI dashboard, teachers who are end users can more easily help their students. Hassanudin *et al*. [4] created a predictive BI dashboard to predict the corrosion rate on pipelines using ANN and MLRA and display the result on a BI dashboard created using hyper text markup language (HTML), cascading style sheet (CSS), and JavaScript (JS). This research produces a predictive BI dashboard that can help corrosion engineers to make effective and accurate decisions so as to avoid reducing the risk of financial losses. Cho *et al*. [5] made a predictive BI dashboard using a hybrid method that combines predictions from an unsupervised machine learning model and a semi-supervised machine learning model to predict the condition of factory machines. The predictive BI dashboard is made so that maintenance can be done before the machine is damaged. By developing the predictive BI dashboard, the factory can improve its service quality due to less downtime. Dagliati *et al*. [6] created a predictive BI dashboard to predict the condition of type 2 diabetes patients. The prediction is made using predictive models created in R and MATLAB. The prediction results from the predictive model are then loaded into a data warehouse. Then the data from the data warehouse is displayed on a dashboard created using HTML, CSS, JS, and Google Charts. The dashboard that has been made has proven to have a positive impact where type 2 diabetes patients can reduce the duration of the medical check-ups they need to do [6].

There are shortcomings from the previous research. The predictive BI dashboard implementation is made for businesses which products do not expire easily and the product's trend does not change quickly, unlike the products in the food and beverage businesses. Therefore, the previous research is not suitable to be implemented in the food and beverage business. This study was made to overcome the shortcomings of previous research by creating a predictive BI dashboard for the food and beverage business using a dataset from a bakery. Predictions that are made in this research are customer segmentation using RFM analysis and MBKM; and the amount of bread that must be produced (demand forecasting) per day using XGBoost. Customer segmentation is done to categorize customers based on their purchasing patterns. Meanwhile, demand forecasting is done to reduce the amount of unsold bread. By identifying the existing customer's behavior and the customer's daily demand, company owners can make decisions more effectively and efficiently.

2.2. Demand forecasting using machine learning

Demand forecasting is the process of using historical data of a company to predict the number of goods that the company must produce in the future. Demand forecasting can provide many advantages for companies such as ensuring that the goods produced always meet the market demand, reducing downtime cost, lowering product costs by implementing a better storage system, and reducing the number of unsold products [8], [9]. Machine learning is often used to perform demand forecasting because it provides good predictive accuracy [10]. There are several research about demand forecasting using machine learning. Tanizaki *et al*. [11] performed demand forecasting on the number of restaurant orders using Bayesian linear regression which produces a model with a model with an accuracy of 80.51%. Mouatadid and Adamowski [12] performed demand forecasting on water consumption using extreme learning machine which produces a model with an accuracy of 81.73%. Moroff *et al*. [13] performed demand forecasting for goods in retail stores using multilayer Perceptron which produces a model with an accuracy of 81.9%. Eseye *et al*. [14] performed demand forecasting on the amount of electricity consumption per hour using feed forward ANN which produces a model with an accuracy of 83%. Abbasi *et al.* [15] conducted research on demand forecasting on electrical load using XGBoost and the predictions made by XGBoost have an accuracy of 97.21%. Sukarsa *et al*. [16] performed demand forecasting on gourami fish supplies using XGBoost and produce a model with an accuracy of 97.54%. Because the accuracy of XGBoost is higher than other methods, XGBoost was chosen as the algorithm to perform demand forecasting in this research.

2.3. Recency, frequency, and monetary value analysis using machine learning

RFM analysis is a method for classifying customers based on how recent a customer made a transaction (recency), how often a customer makes transactions (frequency), and how much a customer usually spends per transaction (monetary value) [17], [18]. RFM analysis can be used to help a business market its products by finding each customer's buying pattern and classifying customers who have similar buying pattern [19]. By performing RFM analysis, businesses can tailor their marketing campaigns to each customer group, so that marketing campaigns can better suit the needs of each customer group [20]. In RFM analysis, a customer can be categorized as a loyal or good customer if the customer has a high RFM [21]. RFM analysis is performed by calculating the RFM for each customer. After that, RFM analysis will divide customers into several categories so that the company can provide personalized services according to the behavior of the customer group [22].

After conducting a literature review, several research on RFM analysis methods were found. Rojlertjanya performs RFM analysis using transactional data of IT companies in Thailand. The research was conducted by developing a k-means model and the model produced a silhouette coefficient (SC) of 0.4 [23]. Then, research by Shirole *et al*. [24] perform RFM analysis using k-means and UK's e-commerce dataset which produces a model with an SC of 0.44. Furthermore, research by Kara [25] did an RFM analysis using k-means and transactional data from an electronics company in Istanbul which resulted in a model with an SC of 0.51. Gustriansyah *et al*. [26] perform RFM analysis using k-means and transactional data from pharmacies in Indonesia which produces a model with an SC of 0.52. Based on the literature review, it was found that k-means is the most frequently used algorithm for RFM analysis. Therefore, in this research, a variation of k-means algorithm will be used, namely MBKM algorithm. MBKM is used in this research because MBKM works the same way with k-means, but the data used is split into batches, thus providing parallelism capabilities that can speed up the clustering process.

3. OVERVIEW OF PREDICTIVE BI DASHBOARD

This section will describe the components and technologies used in the predictive BI dashboard. Then, a more detailed description of the machine learning models development is explained in the following two sub sections: subsection 3.1 describe the development processes of the XGBoost model and subsection 3.2 describe the development processes of the MBKM model. The predictive BI dashboard's components can be seen through Figure 1.

Based on Figure 1, XGBoost algorithm is used to predict the amount of bread that must be produced by the bakery. The MBKM algorithm is used to perform clustering on the RFM data from each customer. The results of the XGBoost and MBKM algorithms are then inserted into a staging database. The staging database also contains transactional data belonging to the bakery business. After the data is inserted into the staging database, the extract, transform, and load (ETL) process is performed by transforming the staging database's data to fit the data warehouse's structure using an ETL script. The data warehouse created for the bakery business was developed using the bottom-up approach (Kimball approach). The data in the data warehouse are displayed using a BI dashboard so that end users can see the prediction results for the amount of bread that must be produced, the customers's category from the RFM analysis, and the bakery's performance from transactional data. In the implementation of this research, several technologies and libraries are used: "XGBRegressor" function from the "xgboost" library is used to create an XGBoost model; “MiniBatchKMeans” function from the “sklearn” library is used to create a MBKM model; MySQL is used to create the staging database; PostgreSQL is used to create the data warehouse; Python programming language, “numpy” library, and “pandas” library are used to create ETL scripts; and Microsoft Power BI is used to create a BI dashboard.

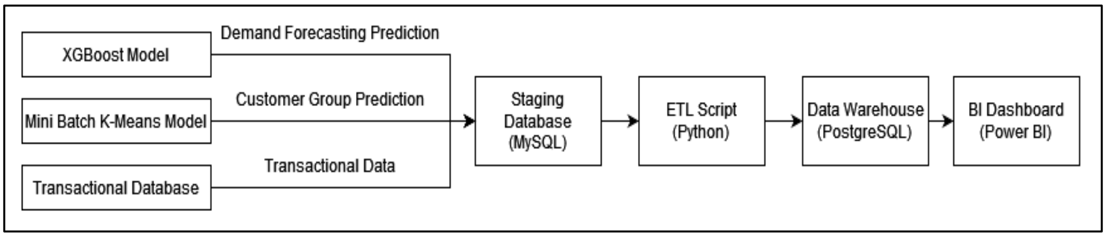

Figure 1. Predictive BI dashboard components

3.1. Extreme gradient boosting model development

The XGBoost model created follows the XGBoost model developed by [27]. The XGBoost model uses transactional data from January 2005 to July 2022. The data preprocessing carried out are: deleting null or NaN rows, performing "JOIN" operations between each table, and deleting unused columns. After that, a feature selection is performed using the “plot_importance” function from the XGBoost library to find the most influential features to predict the amount of bread that must be produced.

Based on the “plot_importance” function, it is found that the number of breads ordered (quantity), the grand total of a transaction (GrandTotal), month of a transaction (month), year of a transaction (year), and day (day of the month for a transaction) are the features that most influence the XGBoost model’s prediction. After performing feature selection, the dataset is divided into two parts: the training dataset and the testing dataset. The 80% of the dataset became the training dataset and the remaining 20% became the testing dataset. Some of the parameters used by the XGBoost model are: the "learning_rate" parameter of 0.6, the "max_depth" parameter of 5, the "reg_lambda" parameter of 0.9, the "reg_alpha" parameter of 0.1, and the "subsample" parameter of 1. After the XGBoost model is created, the model is evaluated using root mean square error (RMSE) and R^2 score.

3.2. Mini batch k-means model development

The MBKM model was created using transactional data from January 2005 to July 2022. To use MBKM in RFM analysis, data preprocessing is carried out on the dataset used, such as: deleting null or NaN rows, deleting unused columns, performing "GROUP BY" operation on customer ID, creating the “recency” column by subtracting the current date with the maximum value of the transaction date, creating the “frequency” column by performing COUNT operation on the transaction ID, and creating the “monetary value” column by performing SUM operation on the grand total of each transaction. After performing the data preprocessing step, the dataset will only have four columns: the "customer id" column, the "recency" column, the "frequency" column, and the "monetary value" column. After the dataset is preprocessed, the dataset can be used by the MBKM algorithm. Some of the parameters used by the MBKM model in this research are: the "n_clusters" parameter of 6, the "max_iter " parameter of 30, the "tol " parameter of 0, the "max_no_improvement" parameter of 0, the "init_size" parameter of 140, the "n_init " parameter of 3, the "reassignment_ratio" parameter of 0.01, the "batch_size" parameter of 1,536, and the "init " parameter of “kmeans++”. After the MBKM model is created, the model is evaluated using Dunn index, SC, and Davies-Bouldin index.

4. METHOD

4.1. Requirement and data gathering

The first development stage includes gathering requirements from end users, collecting data that will be used for the data warehouse, and conducting a literature review to find out the best method to use in this research. The data obtained from the end user comes from the bakery’s transactional database in the form of comma separated value (CSV) files. The transactional database has six tables: the "area" table which contains the areas of the bakery’s customers; the "customer" table containing information about the bakery's customers; the "inventory" table containing information about the products sold at the bakery; the "sales_header" table containing the transaction time, transaction date, name of the shop that performs the transaction, and the total transaction amount; the "sales_detail" table which contains information about the products purchased, the quantity, and the sub total of the product.

4.2. Data warehouse development

In the second stage, the data warehouse is developed using the nine steps of data warehouse design method created by Kimball and Ross [28], [29]. The data warehouse's star schema can be seen in Figure 2. Based on Figure 2, it can be seen that there is one fact table and eight dimension tables. The fact table or "sales_detail_fact" contains 25 columns, but those 25 columns can be categorized into six types of columns: composite key columns, alternate key columns, foreign key columns, columns that contains data from the transactional database, columns that contains the demand forecasting result from XGBoost, and columns that contains result from RFM analysis using MBKM. As for the dimension table, there are eight dimension tables: the "product_dim" table which contains the product name; the "customer_dim" table containing customer's name and customer's status (active or inactive); the "date_dim" table containing transaction date, year, quarter, month, week of the year, and day of the year; the "sales_person_dim" table containing the name of the salesperson in charge of the transactions; the "area_dim" table containing area name, latitude, and longitude; the "sales_type_dim" table containing the type of transaction (cash or credit); the "sales_approval_dim" table which contains the approval status of a transaction (approved or have not been approved); and the table "rfm_segmentation_dim" table which contains the name of the customer group resulting from the RFM analysis.

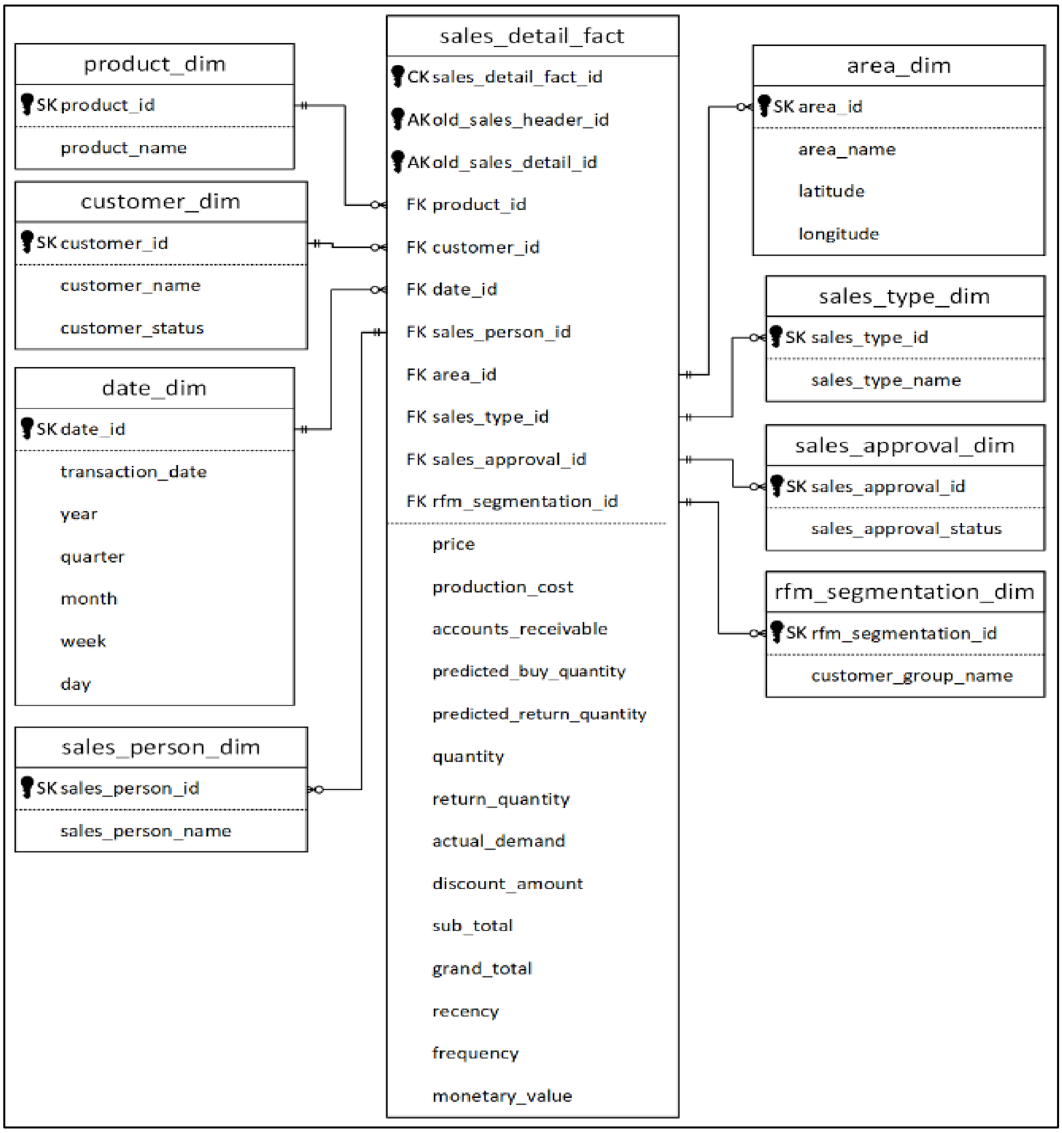

Figure 2. Bakery data warehouse

4.3. Extract, transform, and load

In the third stage, the demand forecasting result from XGBoost and the customer segmentation result from the MBKM will be combined with data from the transactional database and loaded into a staging database. After that, the data from each staging database table will be converted into CSV so that it can be processed using an ETL script written in Python. The first step in the ETL script is to read the CSV of all the staging database's table and convert it to pandas dataframes. The second step in the ETL script is to perform various transformations to the pandas dataframes, such as: deleting unused columns, performing a "JOIN" query with several table dataframes, creating surrogate keys, creating composite keys, creating new date columns, and sorting the modified columns so that it has the same structure as the data warehouse tables in Figure 2. The third step in the ETL script is to convert the transformed dataframes into CSV files. After the ETL script has been executed, it will generate several CSV files for each data warehouse table. The next process is to load the data from the CSV into the appropriate table in the data warehouse.

4.4. Business intelligence dashboard development

In the fourth stage, a BI dashboard was created using data from the data warehouse. The BI dashboard is divided into two sections: the RFM analysis section and the demand forecasting section. The RFM analysis section consists of five charts regarding RFM analysis: the "number of customers per RFM category" chart which can be used to view the number of customers per RFM category who make transactions per month; the "total revenue per RFM category" chart which can be used to see the total revenue of each RFM category per month; the "average RFM value per RFM category" chart that can be used to see the average RFM of each RFM category; "RFM value per customer" chart that can be used to view the RFM per customer; and the "RFM categorization per area" chart that can be used to compare the number of customers per RFM category in an area. The RFM analysis section can be seen in Figure 3. The demand forecasting section includes: "approval status for the last 14 days" chart which can be used to view the number of transactions that have been approved and have not been approved and the "demand forecasting for the last 14 days" chart that can be used to see XGBoost's predictions and the actual bread demand. The demand forecasting section can be seen in Figure 4.

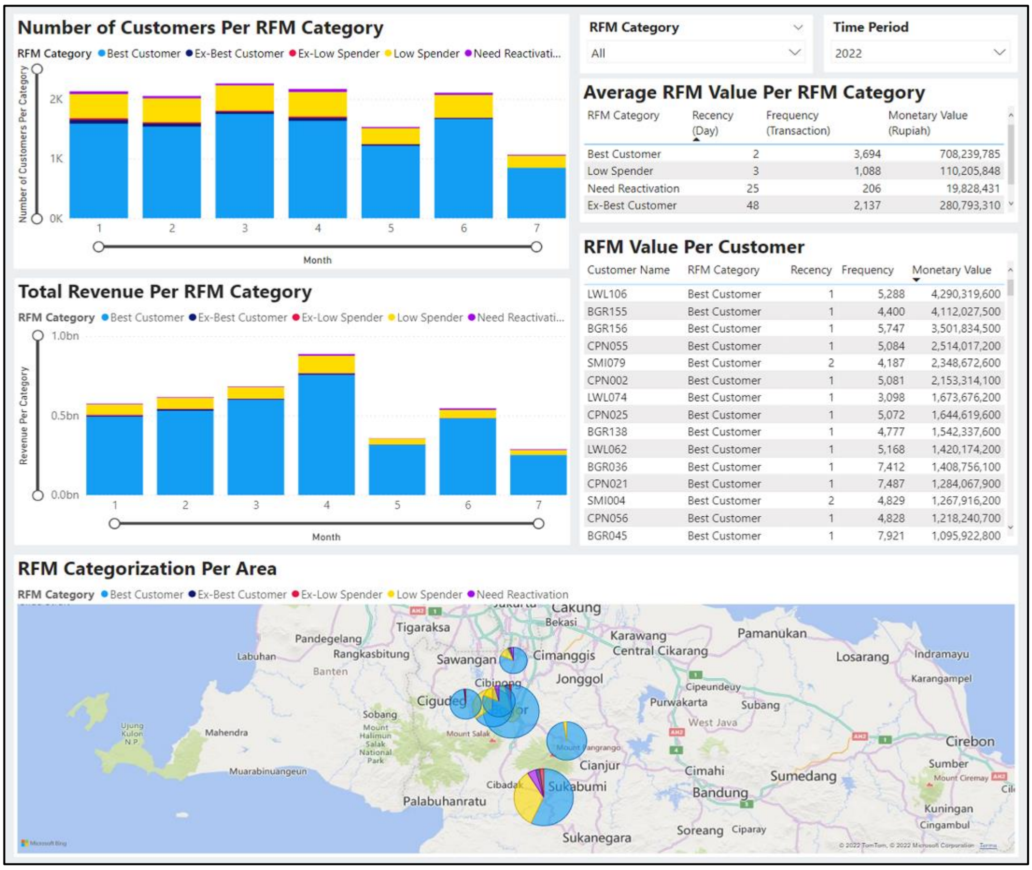

Figure 3. RFM analysis section

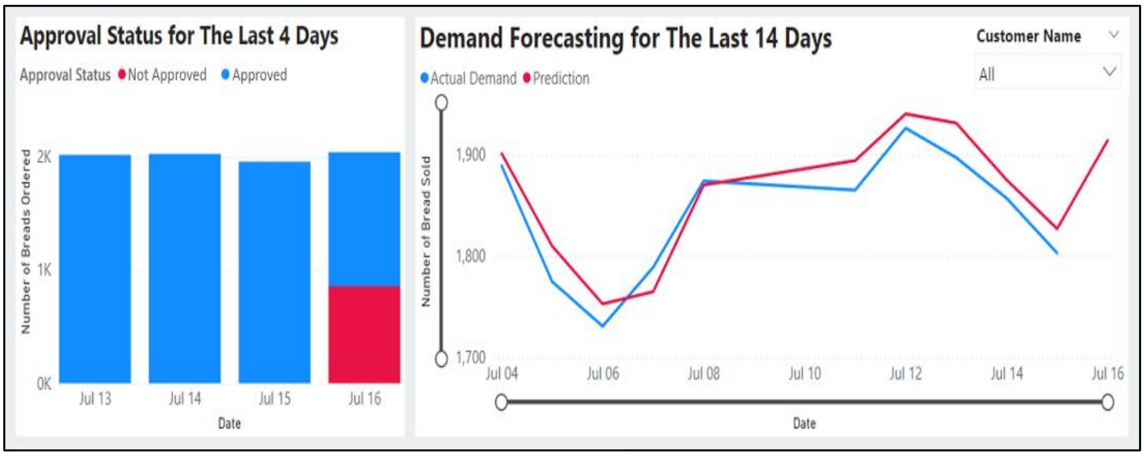

Figure 4. Demand forecasting section

4.5. Machine learning models and business intelligence dashboard evaluation

In the fifth stage, three evaluations were carried out: BI dashboard evaluation, XGBoost model evaluation, and MBKM model evaluation. The XGBoost model is evaluated using RMSE and R^2 score. The MBKM model is evaluated using Dunn index, SC, and Davies-Bouldin index. The BI dashboard were evaluated by end users using a questionnaire based on Ellis's method in evaluating a good application's user interface [30]. The questionnaire contains ten questions, which assess four important aspects: "ease of use", "ease of understanding", "error-free rate", and "BI dashboard's effectiveness for the company's end goal". The four aspects will be assessed using ten questions and each of the question is answered using a number from one to five, where a score of one means that the developed BI dashboard failed to fulfill the assessed aspect, while a score of five means that the developed BI dashboard successfully fulfill the assessed aspect. The score of each question in one aspect will be averaged so that a final score is obtained for the evaluated aspect. This process will be carried out for each aspect, so that four final scores are obtained for the four aspects evaluated. After the evaluation stage has been carried out, the BI dashboard development process is complete. The development stages carried out in subsection 4.1 to subsection 4.5 can be visualized through Figure 5.

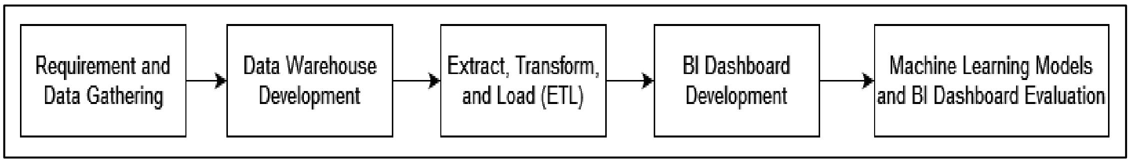

Figure 5. Predictive BI dashboard development stages

5. RESULTS AND DISCUSSION

This section will be divided into three sub sections: subsection 5.1, subsection 5.2, and subsection 5.3. Subsection 5.1 will explain the BI dashboard evaluation process. Subsection 5.2 will explain the XGBoost model evaluation process. Subsection 5.3 will explain the MBKM model evaluation process.

5.1. Extreme gradient boosting model evaluation

To evaluate XGBoost's performance, XGBoost was compared with decision tree, random forest, support vector regression (SVR), and lasso regression algorithms using data from one of the bakery customers who made the most transactions as the training and testing dataset. The dataset from the customer will be split into two parts so that 70% of the data will be used for training the dataset and 30% of the data will be used for testing the dataset. The testing results of the five algorithms can be seen in Table 1. Based on Table 1, XGBoost got the best value in all evaluation metrics used. Therefore, it is proven that XGBoost is the best algorithm to use compared to the other four algorithms.

Table 1. Algorithms evaluation for bakery demand forecasting

Algorithm	Evaluation metric	
	RMSE	R^2 score
XGBoost	0.188	0.931
Decision tree	0.283	0.844
Random forest	0.291	0.834
SVR	0.746	-0.091
Lasso regression	0.747	-0.093

5.2. Mini batch k-means model evaluation

To evaluate the performance of MBKM, the algorithm was compared with agglomerative clustering, balanced iterative reducing and clustering using hierarchies (BIRCH), k-means, spectral clustering, and Gaussian mixture model (GMM). The training and testing data used for the models are the RFM of each customer. The comparison result of the six algorithms can be seen in Table 2.

Based on the data from Table 2 k-means, MBKM, and GMM get the highest score from the six algorithms tested. Although the three models have the best performance from the other algorithms, the three models have Dunn index, silhouette score, and Davies-Bouldin index values that differ slightly from each other. Although the performance of the three models is similar, MBKM is considered the best algorithm to perform RFM analysis compared to k-means and GMM. The reason is explained in a research conducted by Kubara where although GMM has the shortest code run time, it easily get stuck in the local minimum; k-means itself has a drawback where the code run time is slower than GMM even though it has better clustering performance than GMM [31]. MBKM does not have these two problems because the MBKM is not easily trapped in the local minimum because it uses the same algorithm as k-means, but the MBKM is faster than k-means to run the code because it can perform parallelism. Therefore, it is proven that MBKM is the best algorithm to perform RFM analysis.

From the evaluation result and hyperparameter tuning, it was also found that the most suitable number of clusters for the bakery data was 6. Then each of the six clusters is given a category name: “best customer”, “low spender”, “need reactivation”, “ex-best customer”, “ex-low spender”, and “lost customer”. The names are given according to the behavior of each cluster which can be seen from the average RFM. The following is the average RFM of the six clusters.

Based on data from Table 3, each customer category has behavior that distinguishes it from other categories. Customers in the “best customer” category are those who have the best RFM. Customers in the “best customer” category is the most profitable customer for the business. Customers in the "low spender" category are customers whose frequency and monetary value are much smaller than the "best customer", but the "low spender" category has almost the same recency value as the "best customer" category. This means that this customer is a customer who has just shopped, but the frequency of shopping and the nominal of each transaction is small. Customers with the "need reactivation" category are customers whose recency and frequency values, and their monetary value are lower than other categories, so it is assumed that customers in this category are no longer shopping at the bakery. The name “need reactivation” is used to indicate that customers who are in this category need to be visited immediately before the customer moves to another competitor. Customers with the "ex-best customer" category are customers who have a frequency and monetary value similar to "best customer", but the customer's recency value is high which means the customer has not shopped at the bakery for a long time. Customers in the "ex-low spender" category are customers who have a frequency and monetary value similar to "low spender", but the customer's recency value is high as in the "ex-best customer" category, which means that the customer has not shopped at the bakery for a long time. Customers in the "lost customer" category are customers whose recency value is so large that they are considered as customers who have moved to other competitors or whose business has closed. Customers who fall into this category are customers who have not shopped at the bakery for two years.

Table 2. Algorithms evaluation for RFM analysis

Algorithm	Evaluation metric		
	Dunn index	Silhouette score	Davies-Bouldin index
MBKM	0.4264	0.4421	0.8327
Agglomerative clustering	0.3333	0.4026	0.8816
BIRCH	0.3779	0.4035	0.8896
K-means	0.4264	0.4439	0.8441
Spectral clustering	0.4264	0.4411	0.8622
GMM	0.4264	0.4420	0.8280

Table 3. Average RFM value for each cluster

RFM category	Recency (day)	Frequency (transaction)	Monetary value (*rupiah*)
Best customer	2	3,694	708,239,785
Low spender	3	1,088	110,205,848
Need reactivation	25	206	19,828,431
Ex-best customer	48	2,137	280,793,310
Ex-low spender	52	625	56,610,422
Lost customer	4,994	178	7,158,648

5.3. Business intelligence dashboard evaluation

From the BI dashboard that was made, analysis that can be done is to use the "demand forecasting for the last 14 days" chart and the "approval status for the last 14 days" chart to consider the amount of bread that must be produced and compare it with the actual demand from the previous day. Then, end users can use the "number of customers per RFM category" chart and the "total revenue per RFM category" chart to prioritize marketing campaigns for the RFM category with the most members and the RFM category that provides the most revenue for the company. The next analysis that can be done is to use the "average RFM value per RFM category" chart and "RFM value per customer" chart to compare the RFM value of each customer with the average RFM value of that category. From this comparison, end users can see customers whose RFM value is much worse than the average RFM value of that category, so the company can quickly act to these customers such as by giving discounts or creating a bundling program to increase the RFM value of these customers. Then, the "RFM categorization per area" chart can be used by the factory manager to evaluate the performance of the salesperson who manages the area, which can be done by analyzing the customer's category changes in one of the areas. As described in subsection 4.5. the BI dashboard is evaluated by the end user using a questionnaire. The Figure 6 is the evaluation score given by the end-user to the BI dashboard that was created.

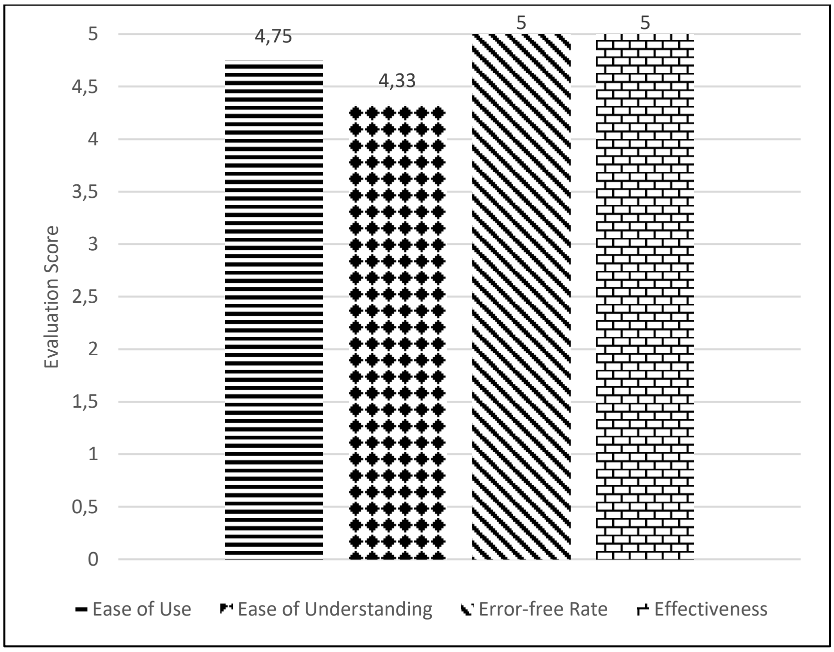

Figure 6. BI dashboard evaluation questionnaire score

Based on Figure 6, the average value for the "ease of use" aspect is 4.75 out of 5, the average value for the "ease of understanding" aspect is 4.33 out of 5, the average value for the "error-free rate" aspect is 5 out of 5, and the average score for the "BI dashboard's effectiveness" aspect is 5 out of 5. If these values are averaged, the BI dashboard produces a final score of 4.77 out of 5. The end user also described that the BI dashboard provides three benefits: end users can speed up decision-making process because the end users do not need to make company performance reports which usually take four hours on a regular basis, the bakery business can find new insights that previously could not be seen without data visualization, and decision making becomes more accurate because it is based on the actual data.

6. CONCLUSION

This research has succeeded in developing a predictive BI dashboard that can help analyze the food and beverage businesses (businesses that sell fast-expired goods). The developed predictive BI dashboard has succeeded in accelerating the decision-making process and has succeeded in helping bakery businesses discover new insights. In addition, based on the evaluation results, the XGBoost and MBKM models that were created succeeded in providing good and accurate prediction results. There are several improvements that can be made in future research, such as automating the ETL process, experimenting with other demand forecasting methods and algorithms to improve the prediction accuracy, and experimenting with other customer segmentation methods and algorithms to improve the clustering accuracy.

REFERENCES

[1] A. Grabińska and L. Ziora, "The application of business intelligence systems in logistics. Review of selected practical examples," *System Safety: Human - Technical Facility - Environment*, vol. 1, no. 1, pp. 1028–1035, Mar. 2019, doi: 10.2478/czoto-2019-0130.

[2] B. Hansoti, "Business intelligence dashboard in decision making," M.S. thesis, Dept., Technol., Purdue Univ., West Lafayette, IN, USA, 2010.

[3] N. Diana, M. Eagle, J. Stamper, S. Grover, M. Bienkowski, and S. Basu, "An instructor dashboard for real-time analytics in interactive programming assignments," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Mar. 2017, pp. 272–279, doi: 10.1145/3027385.3027441.

[4] S. N. Hassanudin, I. A. Aziz, J. Jaafar, S. Qaiyum, and W. M. A. M. Zubir, "Predictive analytic dashboard for desalter and crude distillation unit," in *2017 IEEE Conference on Big Data and Analytics (ICBDA)*, Nov. 2017, pp. 55–60, doi: 10.1109/ICBDAA.2017.8284107.

[5] S. Cho *et al.*, "A hybrid machine learning approach for predictive maintenance in smart factories of the future," in *IFIP Advances in Information and Communication Technology*, 2018, pp. 311–317, doi: 10.1007/978-3-319-99707-0_39.

[6] A. Dagliati *et al.*, "A dashboard-based system for supporting diabetes care," *Journal of the American Medical Informatics Association*, vol. 25, no. 5, pp. 538–547, May 2018, doi: 10.1093/jamia/ocx159.

[7] S. Palanisamy, "Predictive analytics with data visualization," *Journal of Ubiquitous Computing and Communication Technologies*, vol. 4, no. 2, pp. 75–96, Jul. 2022, doi: 10.36548/jucct.2022.2.003.

[8] J. J. Bergman, J. S. Noble, R. G. McGarvey, and R. L. Bradley, "A Bayesian approach to demand forecasting for new equipment programs," *Robotics and Computer-Integrated Manufacturing*, vol. 47, pp. 17–21, Oct. 2017, doi: 10.1016/j.rcim.2016.12.010.

[9] J.-H. Böse *et al.*, "Probabilistic demand forecasting at scale," *Proceedings of the VLDB Endowment*, vol. 10, no. 12, pp. 1694–1705, Aug. 2017, doi: 10.14778/3137765.3137775.

[10] K. Kaya, Y. Yılmaz, Y. Yaslan, Ş. G. Öğüdücü, and F. Çıngı, "Demand forecasting model using hotel clustering findings for hospitality industry," *Information Processing & Management*, vol. 59, no. 1, p. 102816, Jan. 2022, doi: 10.1016/j.ipm.2021.102816.

[11] T. Tanizaki, T. Hoshino, T. Shimmura, and T. Takenaka, "Demand forecasting in restaurants using machine learning and statistical analysis," *Procedia CIRP*, vol. 79, pp. 679–683, 2019, doi: 10.1016/j.procir.2019.02.042.

[12] S. Mouatadid and J. Adamowski, "Using extreme learning machines for short-term urban water demand forecasting," *Urban Water Journal*, vol. 14, no. 6, pp. 630–638, Jul. 2017, doi: 10.1080/1573062X.2016.1236133.

[13] N. U. Moroff, E. Kurt, and J. Kamphues, "Machine learning and statistics: a study for assessing innovative demand forecasting models," *Procedia Computer Science*, vol. 180, pp. 40–49, 2021, doi: 10.1016/j.procs.2021.01.127.

[14] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems," *IEEE Access*, vol. 7, pp. 91463–91475, 2019, doi: 10.1109/ACCESS.2019.2924685.

[15] R. A. Abbasi, N. Javaid, M. N. J. Ghuman, Z. A. Khan, S. U. Rehman, and Amanullah, "Short term load forecasting using XGBoost," in *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019)*, 2019, pp. 1120–1131, doi: 10.1007/978-3-030-15035-8_108.

[16] I. M. Sukarsa, N. N. P. Pinata, N. K. D. Rusjayanthi, and N. W. Wisswani, "Estimation of gourami supplies using gradient boosting decision tree method of XGBoost," *TEM Journal*, vol. 10, no. 1, pp. 144–151, Feb. 2021, doi: 10.18421/TEM101-17.

[17] J. Wu and Z. Lin, "Research on customer segmentation model by clustering," in *Proceedings of the 7th international conference on Electronic commerce - ICEC '05*, Jun. 2005, pp. 316–318, doi: 10.1145/1089551.1089610.

[18] O. Dogan, B. Oztaysi, and A. Isik, "Fuzzy RFM analysis in car rental sector," *International Journal of Technology and Engineering Studies*, vol. 7, no. 2, pp. 8–14, 2021, doi: 10.20469/ijtes.7.10002-2.

[19] T. Juhari and A. Juarna, "Implementation RFM analysis model for customer segmentation using the k-means algorithm case study XYZ bookstore," *EXPLORE*, vol. 12, no. 1, pp. 107–118, 2022.

[20] R. C. Blattberg, B.-D. Kim, and S. A. Neslin, *Database marketing: analyzing and managing customers*, 1st ed. New York: Springer, 2008.

[21] S. Wan, J. Chen, Z. Qi, W. Gan, and L. Tang, "Fast RFM model for customer segmentation," in *Companion Proceedings of the Web Conference 2022*, Apr. 2022, pp. 965–972, doi: 10.1145/3487553.3524707.

[22] H.-C. Chang and H.-P. Tsai, "Group RFM analysis as a novel framework to discover better customer consumption behavior," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14499–14513, Nov. 2011, doi: 10.1016/j.eswa.2011.05.034.

[23] P. Rojlertjanya, "Customer segmentation based on the RFM analysis model using k-means clustering technique: a case of IT solution and service provider in Thailand," M.S. thesis, Dept. Manage. (Bus. Innov.), Bangkok Univ., Pathum Thani, Thailand, 2019.

[24] R. Shirole, L. Salokhe, and S. Jadhav, "Customer segmentation using RFM model and k-meansclustering," *International Journal of Scientific Research in Science and Technology*, vol. 8, no. 3, pp. 591–597, Jun. 2021, doi: 10.32628/IJSRST2183118.

[25] O. F. Kara, "Customer clustring with machine learning," M.S. thesis, Dept. Inf. Technol., MEF Univ., İstanbul, Türkiye, 2021.

[26] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis based on k-means," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 470–477, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp470-477.

[27] N. Wikamulia, M. Jonathan, and S. M. Isa, "Bakery demand forecasting using Xgboost and k-means clustering," *ICIC Express*

Letters, Part B: Applications, vol. 14, no. 1, pp. 21–28, 2023, doi: 10.24507/icicelb.14.01.21.
[28] R. Kimball and M. Ross, "The data warehouse toolkit: the complete guide to dimensional modeling," pp. 101–102, Apr. 2002.
[29] R. Kimball and M. Ross, *The Kimball group reader: relentlessly practical tools for data warehousing and business intelligence*. Indianapolis: Wiley Publishing, 2010.
[30] T. Ellis, "User interface: good design vs bad design," *H2O Digital*, 2017. https://h2o-digital.com/user-interface-good-design-vs-bad-design/ (accessed May 14, 2022).
[31] K. Kubara, "Gaussian mixture models vs k-means. Which one to choose?," *Towards Data Science*, 2020. https://towardsdatascience.com/gaussian-mixture-models-vs-k-means-which-one-to-choose-62f2736025f0 (accessed Apr. 26, 2022).

BIOGRAPHIES OF AUTHORS

Nathaniel Wikamulia received a Bachelor of Computer Science from Bina Nusantara University in 2022 and currently pursuing a master's degree in computer science from Bina Nusantara University. His research interests are in the areas of business intelligence, machine learning, and data science. He can be contacted at email: nathaniel.wikamulia@binus.ac.id.

Sani Muhamad Isa completed his Bachelor of Mathematics from Padjadjaran University in 1995, his master's in computer science from University of Indonesia in 2001 and, Ph.D. in Computer Science from University of Indonesia in 2014. He is currently serving as lecturer and Director of Graduate Program in Bina Nusantara University. His research interests are in the areas of biomedical engineering, remote sensing, and data science. He can be contacted at email: sani.m.isa@binus.ac.id.