

Implementation of Business intelligence in Know Your Customer (KYC) for credit card customers' loan repayment status

Mario Caesar Kristantoputra^{1*} and Preethi Subramanian¹

¹Asia Pacific University, Jalan Teknologi 5, Taman Teknologi Malaysia, 57000 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

Abstract. Nowadays, the increasing technological improvements followed by data demand made businesses and organizations follow these improvements to continue operating. The Know Your Customer (KYC) assessment is still manually done, followed by the growing amount of data, resulting in data accumulation and affecting the time organizations spend analyzing the prospective applicant data. The unavailability of tools that can help review documents and data leads to more data accumulation and time-consuming KYC assessments. This paper aims to create a Business Intelligence (BI) system to help financial organizations analyze, process applicant data, and determine applicants' eligibility who are willing to get credit card services. This paper may be utilized by an enterprise operating in the financial sector that follows the KYC procedure to identify applicants who require credit card services. The development of BI systems is predicted to help reduce the time spent to validate application data, particularly in the banking or financial industry. This study has designed three dimension tables and a fact table using Microsoft SQL Server 18 for the data warehouse. Pentaho Data Integration is used for the ETL process, and Tableau creates the dashboard. The dashboard contains general information and the loan repayment status of an applicant. Two pivot tables were created using Microsoft Excel to summarize the loan repayment status of an applicant. **Keywords:** Business Intelligence; Know Your Customer; Loan Repayment; Credit Card; ETL Process; Tableau.

1 Introduction

The need for data is growing due to technological advancements, where businesses or organizations need to continue operating. Data is critical for companies and organizations since it can be used to make decisions and determine an organization's/business company's strategy. Due to the increased need for data, many businesses and industries are now implementing technology capable of processing and analyzing data to provide knowledge or insights that can be used for the business purposes of the company or organization. The creation of BI is one of the ways companies provide for gathering essential insights from data (Prasad, 2021). Organizations may use business solutions to address issues promptly and

* Corresponding author: tp061966@mail.apu.edu.my

adequately and increase operational efficiency while creating a report. Banks can more readily comprehend data using business intelligence tools such as graphs, charts, and animation (Hertz, 2017).

Al Baraka Bank Sudan is an example of how the BI system is used in banking. Al Baraka Bank employs a BI system called iMAL BI core, developed by Path Solutions (Finextra, 2020). Al Baraka Bank Sudan benefits from the use of BI. The benefits include assisting the bank in acquiring and evaluating data and increasing operational productivity by deploying a new end-to-end data warehousing, automated ETL process, and a progressive data model, which leads to effective and efficient decision-making (Finextra, 2020; Hamilton, 2020). Obtaining, categorizing, evaluating, and recognizing trends in data helps Al Baraka Bank Sudan better understand its consumers, increase its ability to forecast sales growth, and mitigate business risks.

Step (2017) and Hoffman (2019) in their studies found that the systematic process of checking prospective application data and the difficulty in establishing the information's legitimacy still make the KYC process onerous. As a result, there is an accumulation of accounts opened by new credit card applicants. In addition, according to research conducted by Trulioo (2019), the KYC system in the banking industry requires appropriate time and resources for bank employees to learn and comprehend KYC norms and regulations. This suggests that no tools are available to analyze credit card applicant data.

Additionally, KYC processes used by financial institutions, particularly banks, currently need a significant amount of time to check an application potential. This issue occurred due to the company's requirement to check and verify the physical document submitted by the candidate with the digital files previously entered into the system. According to Syah et al. (2020), potential users of e-money applications must wait 72 hours before accessing certain features, such as recharging balances, purchasing things from specific stores or businesses, reviewing past transactions, and personal financial planning. These are important since the e-money app company must confirm the candidate's identity by carefully studying previously submitted documents. The e-money provider must also ascertain whether the applicant's documents are genuine.

The BI system developed in this research paper will facilitate banks in improving the KYC procedure and evaluating credit card information for people seeking credit card services. The process of discovering and evaluating possible application data may be done more efficiently and accurately in assessing the feasibility of a credit card service by leveraging data warehouse technology. Additionally, the ETL process will be conducted on the potential applicant's data and the development of a dashboard and pivot table to support banks in making quick and effective eligibility assessments. This study paper may be utilized by a company that runs in the financial industry and complies with KYC rules.

2 Literature Review

This section will perform a literature review on BI adoption in banks. Furthermore, this literature review analyzed the BI system, and the utilization of dashboards may aid banks in processing application data. Additionally, this section will discuss past studies to discover the distinctions between one study and another.

In the previous research by Husni & Mukhlash (2014), BI is utilized to evaluate corporate growth, namely the fundraising and loan success of each bank's XYZ services, branches, and regions. The BI installation began with the design of the star schema, the structure of the generated report, and the cube development. In addition, BI deliverables in Husni & Mukhlash's (2014) study included a monthly company growth report generated using SAS Enterprise Guide.

In addition, according to a study by Owusu et al. (2017), applying BI in financial services might assist institutions in analyzing customer behavior to spot fraud or suspicious activities. As previously mentioned, Al Baraka Bank Sudan has also chosen the business intelligence (BI) system iMAL*BI (IBS Intelligence, 2020). Al Baraka Bank Sudan's dashboards enable companies to monitor each bank product's real-time performance using various In addition, the built BI system helped banks make business decisions.

According to Prayitno's (2018) research, the BI implementation allowed bank ABC to monitor its existing banking products' performance easily and formulate the optimal business strategy. In addition, utilizing the BI system might provide dynamic and visually appealing reports and dashboards.

3 Methodology/Materials

This section will discuss the datasets used and the research methodology of this study, including OLTP development, ETL process, data warehouse design, and OLAP analysis. OLAP analysis will take two steps: pivot table creation and data visualization development in a dashboard.

3.1 Metadata

The datasets for this research were collected from Seanny (2020) via Kaggle, a publicly available dataset resource. This study consists of 2 datasets. The first dataset includes the applicant's data, such as personal information, held property, shares, etc. The following dataset contains all payment/default information linked with a specific customer. The variables included in the datasets are listed in Table 2.

Table 1. Datasets Metadata.

Attributes	Description
application_record.csv	
ID	Identification number for the applicant
CODE_GENDER	Applicant gender
FLAG_OWN_CAR	The applicant owns a car
FLAG_OWN_REALTY	The applicant owns a property
CNT_CHILDREN	Total number of children
NAME_FAMILY_STAT_US	Applicant marital status
NAME_HOUSING_TYPE	House type of an applicant
DAYS_BIRTH	Applicant birthday
DAYS_EMPLOYED	Date of employment start (Count in reverse order from today's date)
FLAG_MOBIL	Mobile phone the applicant owns
FLAG_WORK_PHONE	The applicant owns a work phone
FLAG_PHONE	The applicant owns a phone
FLAG_EMAIL	The applicant owns an email
OCCUPATION_TYPE	Applicant occupation type
CNT_FAM_MEMBERS	Family size
credit_record.csv	
ID	Identification number for the applicant
MONTHS_BALANCE	Record month (The extracted data month acts as the starting point for the subsequent months)

STATUS	Loan status
--------	-------------

3.2 Research Methodology

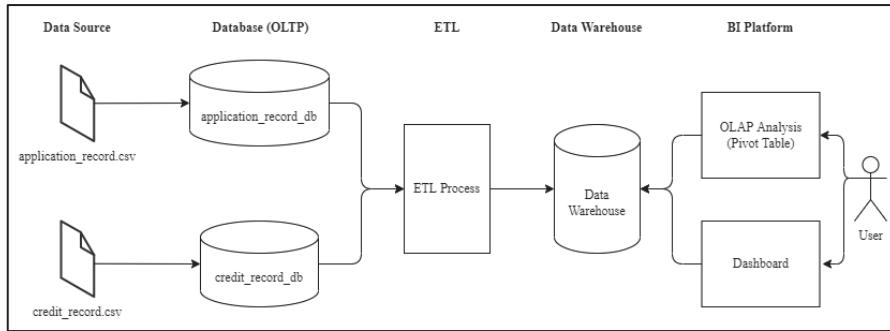


Fig. 1. Development Framework for BI Systems.

The framework for the BI system is represented in Figure 1. The framework is divided into five stages: source of data, OLTP development, ETL process, data warehouse, and BI platform. The first step is OLTP development, which stores information from both datasets into databases before going to the ETL process. All datasets were loaded using Microsoft SQL Server Management Studio 18 to create both databases.

The ETL Process is the subsequent stage. Pentaho Data Integration will be used for the ETL process (also known as Kettle). Certain data pre-processing and transformation steps will be conducted during the ETL process to convert raw data to an appropriate/readable format. The following stage is to design the data warehouse. The data warehouse for this research will be constructed utilizing Kimball & Ross’s (2013) nine-step methodology. After the data has been processed through the ETL process, it will be merged and stored in the data warehouse. Storing processed data in a data warehouse may be accessible in a dashboard and pivot table form to analyze application data for credit card facilities. The data warehouse will include three dimension tables and one fact table. The applicant dimension table is used to store processed data from “application_record_db”) while credit record dimension table (used to store processed data from “credit_record_db”). The time dimension table records the time and date when the ETL process is done running. The single fact table, notably credit card fact, was used to maintain an applicant’s final loan status together with the applicant’s ID. The data warehouse’s Star Schema is seen in Figure 2.

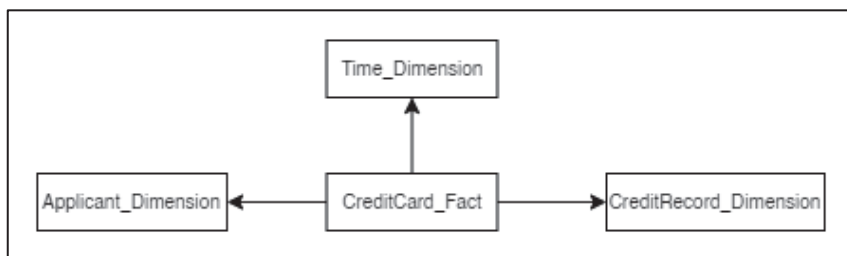


Fig. 2. Design of Star Schema.

After the data is entered into the data warehouse, the pivot table and dashboard are produced. There are two pivot tables. The first pivot table will show information and detail about an applicant with its total good and bad debt. The second pivot table provides

information about each applicant's credit record, including the time and month of each credit record with applicant ID. Microsoft Excel will be used to build all pivot tables. The dashboard displays information about an applicant, such as their applicant ID, personal revenue, gender, property, vehicle ownership, civil status, and type of house.

Furthermore, this dashboard will provide an overview of loan status every month by computing an applicant's status, information, and totals for every loan status. Tableau 2021.1.4 will be utilized to create a dashboard. Figure 3 below shows the dashboard that has been built.

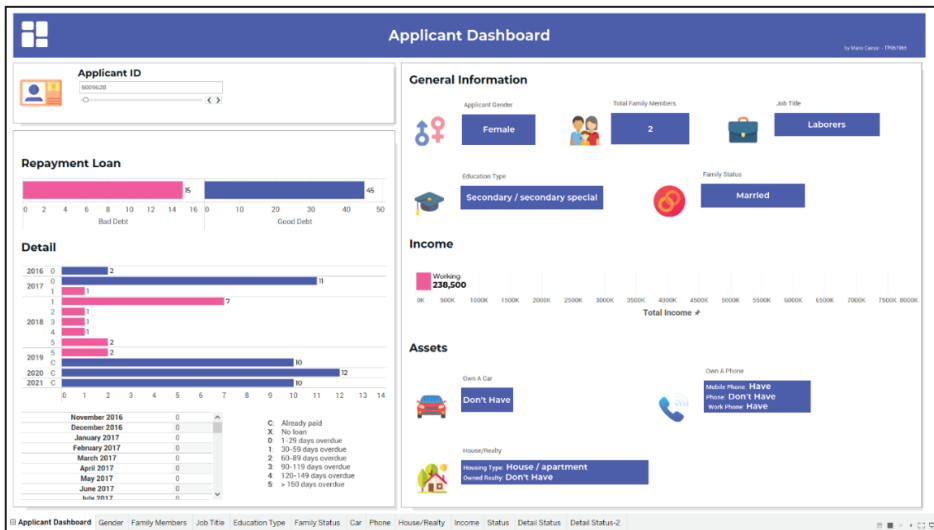


Fig. 3. Applicant Dashboard.

4 Results and Comparison to Previous works

4.1 Results and Findings

Based on the analysis and implementation in the last part, the author may conclude the following results:

- The data warehouse involves three dimension tables: the time dimension, the application dimension, and the credit record dimension, as well as a fact table, credit card fact. The original value/data in the application dimension table has been changed to make the table more straightforward for users to comprehend. Additional columns are provided to the credit record dimension table to indicate an applicant's monthly debt status. Moreover, this dimension table measures an applicant's good and bad status. Time and date variables were given to both tables during the ETL phase.

- The built dashboard provides an applicant's personal information and credit history details.

- Two pivot tables have been created to help with reporting and analysis. The first pivot table provides good and bad debt amounts with its credit record ID and application ID. At the same time, the second pivot table provides information on every month's credit status, which can be searched using a user ID.

The issue stated in the previous section can be addressed since the dashboards and pivot tables are available. It can help financial institutions determine an individual's eligibility for credit card services.

4.2 Comparison to previous works

This part will compare the findings acquired from the BI implementation described in the preceding chapter with those of past comparable studies to compare the methodologies and results obtained by earlier researchers with those of this study.

Table 2. Comparison of This Study to Previous Related Works.

Study Citation	Literature Title	Results
Husni & Mukhlash (2014)	Implementasi Business Intelligence Pada Manajemen Report Bank XYZ	<ul style="list-style-type: none"> At Bank XYZ, business intelligence aided in the reporting process for financing and lending reports, determining the most popular financial products, and determining branch growth. BI system developments began by establishing the Star Schema, then creating generated report layout, followed by the cube design for financing and lending. Finally, SAS Enterprise Guide was used for data processing and creating the OLAP report layout.
Owusu et al. (2017)	Determinants of Business Intelligence Systems Adoption in Developing Countries: an Empirical Analysis From Ghanaian Banks	<ul style="list-style-type: none"> The use of BI in customer relationship management may aid banks in studying client behavior for purposes such as detecting fraud, suspicious activity analysis, and transaction analysis. Relative advantage, complexity, champion presence, corporate preparedness, and regulatory agency primarily influenced BI systems adoption.
IBS Intelligence (2020)	Al Baraka Bank Sudan approached Path Solutions to transform the bank into an intelligent one with iMAL*BI solution	<ul style="list-style-type: none"> The deployment of iMAL*BI affected the overall performance of the bank. The implementation comprised the creation of 24 dashboards, the establishment of an end-to-end data warehousing system, the automation of the ETL process, and the establishment of a progressive data model. The dashboard was created using Power BI. This solution provides various benefits, including the ability to evaluate profitability and bank performance in real-time online utilizing analytical tools, ratios, equations, and computing procedures. Additionally, it enhanced the bank's capacities, easing top management's decision-making.
Prayitno (2018)	Application of Business Intelligence for Banking Performance Based on Products Analysis	<ul style="list-style-type: none"> BI in bank ABC was used to assess and report bank products' performance. By leveraging BI, bank executives could quickly examine a range of bank products. Additionally, BI might assist company executives in strategic decision-

		<p>making on launching prospective bank products with significant advantages.</p> <ul style="list-style-type: none"> • The dashboard displayed information about the savings balance, deposits, and credit accounts. The development of the dashboard used Cognos V 10.2.1.
This Study (2021)	Implementation of Business Intelligence in Know Your Customer (KYC) for Credit Card Customers' Loan Repayment Status	<ul style="list-style-type: none"> • The Tableau-created dashboard can assist banks in evaluating and evaluating application data. It included general information on the applicant and a loan status report. • The pivot tables developed in Microsoft Excel can facilitate banks in evaluating and processing application data and deciding a credit card applicant's eligibility. • Automated ETL with Pentaho Data Integration enables banks to examine applicant data and display converted data in real time.

5 Conclusion

5.1 Discussions

This section will discuss the developed BI system by collecting feedback and comments from a banking professional, particularly in the credit card business, to ensure banks can utilize it to help with the KYC process. Mr. Tjahyono, regional manager of Bank Capital Surakarta, provides assessments and criticism on the BI system. According to Mr. Tjahyono, the built BI system is suitable for banks to review current banking client data to determine the feasibility of obtaining credit card facilities. Additionally, Mr. Tjahyono claimed that the method might assist banks in assessing the viability of clients seeking credit card facilities based on monthly loan payments.

Due to the confidentiality of banking data, the data utilized in this project comes from Kaggle, which can represent the majority of the information required to verify the potential application. However, it is proposed that it would be preferable for the developed system's perfection if it used actual financial data. Banks require additional information or variables (such as monthly costs, alternative sources of income, the names of relatives/people associated with the client, etc.) to verify potential applicants' eligibility for credit card services. By utilizing actual data, it is possible to simulate how financial data is processed until it is loaded into the data warehouse and shown in the dashboard or pivot table. Real datasets or information may also imitate the data cleansing procedure to prepare a bank's data usage.

The ETL process began with data processing in the OLTP environment and ended with data storage in the data warehouse. However, it is advised that the ETL process be automated to minimize the amount of daily data. Furthermore, it enables real-time application data processing, eliminating the need for applicant data to accumulate for analysis due to the automated procedure.

In summary, Mr. Tjahyono said that this study demonstrated the use and advantages of business intelligence in easing banking activity and determining the viability of banking consumers seeking credit cards.

This research applied the previously defined BI system framework effectively. In this study, a data warehouse containing three dimension tables and one fact table has been constructed. Additionally, the ETL automation process can aid banks in providing altered applicant data and determining the actual applicant status. Adding dashboards and pivot tables can also help users evaluate processed data and decide the applicant's eligibility.

Credit card industry banking professionals have also reviewed this study. In this study, it was determined that the deployment of BI might aid banks in processing raw data and KYC processes to assess the eligibility of applicants for existing credit card facilities. In order to achieve optimal outcomes, it is intended that the developed BI system would utilize actual data and other information to aid the process more precisely.

5.2 Importance and Contribution of Study

This study provides two significant contributions to analyzing data on applicants requesting credit card services under the KYC procedure. The first contribution is the dashboard and pivot table development that banks can use to determine applicants' eligibility.

The second contribution is an ETL procedure that may assist with extracting and transforming application data before entering it into the data warehouse. The ETL process assists in filtering out unsuitable application data (for example, applicants under the age of 21 or insufficient data) that is not stored in the data warehouse. Furthermore, the ETL process can help banks simplify the analysis process. Moreover, the ETL process assists in conforming application data to banking requirements.

This research also included assessments and input from banking industry specialists. According to evaluations and feedback, the BI system established allowed banks to review the current banking customer data to determine the possibility of getting credit card facilities. Additionally, the method might assist banks in assessing the viability of clients seeking credit card services. Moreover, this study addressed the research issues raised in the preceding chapter. Furthermore, this study accomplished the general research goals and research aims.

5.3 Future Recommendations

Based on the critiques and comments provided in the preceding section, it is anticipated that future research will be able to duplicate the real-time ETL process using actual bank data. In addition, actual banking data may aid banks in developing applicable dashboards and pivot tables. In addition, it is predicted that further research will be able to automate the ETL process for real-time data, therefore benefiting banks in reducing data growth.

References

1. *Finextra. Al Baraka Bank Sudan introduces iMAL business intelligence core from Path Solutions* (Finextra, 2020) DOI: <https://www.finextra.com/pressarticle/82441/al-baraka-bank-sudan-introduces-imal-business-intelligence-core-from-path-solutions>
2. A. Hamilton, *Path Solutions deploys business intelligence solution at Al Baraka Bank Sudan* (Fintech Futures, 2020) DOI: <https://www.fintechfutures.com/2020/05/path-solutions-deploys-business-intelligence-solution-at-al-baraka-bank-sudan/>
3. I. Hertz, *How business intelligence is revolutionizing banking* (Bobsguide, 2017) <https://www.bobsguide.com/articles/how-business-intelligence-is-revolutionising-banking/>
4. F. Hoffman, *Why Most KYC Fail* (M2SYS, 2019) DOI: <https://www.m2sys.com/blog/biometric-hardware/why-most-kyc-fail/>

5. Z.N. Husni, I. Mukhlash, *Jurnal Sains Dan Senni Pomits* **3**, 16–21 (2014)
6. *IBS Intelligence. Al Baraka Bank Sudan approached Path Solutions to transform the bank into an intelligent one with iMAL*BI solution* (IBS Intelligence, 2020) DOI: <https://ibsintelligence.com/al-baraka-bank-sudan-approached-path-solutions-to-transform-the-bank-into-an-intelligent-one-with-imalbi-solution/>
7. R. Kimball, M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (Third)* (Wiley Publishing, Inc., 2013)
8. A. Owusu, G.C. Agbemabiese, D.T. Abdurrahman, B.A. Soladoye, *Journal of Internet Banking and Commerce* **22(S8)**, 1–25 (2017) DOI: <http://www.icommerceland.com>
9. P. Prasad, *An introduction to Business Intelligence — The basics of BI Roles, Responsibilities, Processes you need to know* (Medium, 2021) DOI: <https://priyashaprasad9511.medium.com/an-introduction-to-business-intelligence-the-basics-of-bi-roles-responsibilities-processes-you-27a288e9617e>
10. D. Prayitno, *Application of Business Intelligence for Banking Performance Based on Products Analysis* **6**, 554–569 (2018)
11. *Seanny. Credit Card Approval Prediction* (Kaggle, 2020) DOI: https://www.kaggle.com/rikdifos/credit-card-approval-prediction?select=application_record.csv
12. A. Step, *Know Your Customer (KYC) in Financial Institutions: Is it my Right or Duty to introduce KYC Documentation* (Tfageeks, 2017) DOI: <http://tfageeks.com/2017/08/21/know-your-customer-kyc-in-financial-institutions-is-it-my-right-or-duty-to-introduce-kyc-documentation/>
13. R. Syah, M.K. Nasution, I. Husein, *Dynamic Control Financial Supervision (OJK) for Growth Customer Behavior using KYC System* **29(8)**, 1110–1119 (2020)
14. *Trulioo. KYC: 3 Steps to Know Your Customer* (Medium, 2019) DOI: <https://medium.com/the-regtech-hub/kyc-3-steps-to-know-your-customer-f788bb35d232>