Journal of Theoretical and Applied Information Technology
31st May 2022. Vol.100. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# LOST WON OPPORTUNITY PREDICTION IN SALES PIPELINE B2B CRM USING MACHINE LEARNING

**AGUSTIAN MAULANA[1], TOGAR ALAM NAPITUPULU[2]**

[1] Information System Management, Binus Graduate Program, Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia

[2] Information System Management, Binus Graduate Program, Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia

E-mail: [1]agustian.maulana@binus.ac.id, [2]tnapitupulu@binus.edu

**ABSTRACT**

Sales pipeline management is needed by companies to manage opportunities. The B2B sales model business has several stages and a long time before dealing with customers. Even though they have used sales pipeline management, the obstacle faced by the sales team and the company is that it is still difficult to analyze data opportunities that have the potential to be lost or won in the early stages. Undetected potential lost or won opportunities at the outset can lead to a risk that many opportunities are lost and the company cannot estimate the planning of resource requirements if the opportunity won. Optimization of B2B CRM data is one of the processes carried out by companies to analyze potentially lost or won. B2B CRM data can be used as supporting customer portfolio data needed to help the sales team analyze lost or won opportunities. Management of historical sales data and customer portfolios using machine learning can generate predictive data that companies need more quickly. Predictive data is needed to help sales teams and companies analyze probability lost or won opportunities more quickly. The sales pipeline management system in B2B CRM assisted by machine learning with the CRISP-DM methodology can identify probability lost or won opportunities in the early stages. So that the sales team has a better chance of converting a lot of opportunities into won.

**Keywords:** *Opportunity, Sales Pipeline, B2B CRM, CRISP-DM, Machine Learning*

## 1. INTRODUCTION

Customer Relationship Management (CRM) is the most popular term in the business world. The CRM concept is embraced by many industries to build relationships with customers dynamically [1]. CRM has been defined as a comprehensive strategy that enables companies to identify, acquire, and retain customers [2]. The customer cycle and operational performance are the main components that can be used as a measurement tool for the contribution of CRM [3], which can drive B2B and B2C business growth [4].

CRM data is something of value and can analyze factors related to customer satisfaction and loyalty [5] which is identical to customer churn analysis. Optimizing other factors such as the sales pipeline can attract new customers, because the acquisition of new customers is very important, including in the B2B business [6].

Business-to-business (B2B) is a form of the business-to-business transaction conducted between companies [7]. The tender and negotiation process is the determinant of loss or won in the B2B sales pipeline.

Sales pipeline is one of the stages in the B2B sales process business that can be interpreted in a CRM application. Eitle et al [6] explain the stages of the sales pipeline in CRM :

a. Leads, a data collection of new prospective customers collected from marketing activities.
b. Opportunities are the process of engagement with customer prospects who are interested in the product.
c. The deal, determining lost or won opportunity.

In practice, the opportunities stage has sub-stages according to the sales process business in the company. The process will be run after the prospective customer's data changes status to qualified [8].

Digitizing the sales pipeline in CRM provides a step forward in maintaining better data quality [9]. Data mining is one of the processes in machine learning to analyze data. This process can help companies to analyze data and predict future trends [10]. The digitization process cannot be separated from the linkage of big data. The CRM process generates a lot of data, starting from the customer portfolio to historical customer satisfaction. With big data processing, CRM data processing using analysis-based predictive models is expected to improve the performance of a company, both B2B and B2C.

To be able to predict lost or won opportunities, a predictive data analysis process is needed in CRM. According to previous research, the use of data mining with data and variables is effective in CRM allows companies to benefit from long-term sales strategies [11],[12]. Testing CRM data using the Support Vector Machine model, generating sentiment variables that can increase the class prediction of lost or won opportunity [13].

CRISP-DM is the standard model most widely used by several data mining researchers to help design data mining projects in addition to Knowledge Discovery Databases (KDD) and SEMMA [14],[15]. Several studies have obtained good results with the CRISP-DM measurement, such as the results of the evaluation of the prediction of the amount of coffee production using a linear regression algorithm that provides good accuracy [16]. And the classification model in predicting players who enter the NBA all-star category using the random forest algorithm produces an average classification accuracy of 92% [17].

For three decades there is still a lack of research on B2B CRM, especially those that discuss the sales pipeline to predict lost or won opportunities [9]. The number of B2B transactions is relatively small and sometimes there is data noise compared to B2C transactions [9]. Currently, problems arise when the company wins the tender process or deals with prospective customers in the negotiation process. where after the process of determining the winner, the period of time to prepare resources for the project is very long. The second problem is that the company is still difficult to determine sales priorities that must be followed-up. In this study, we take a case study on a company engaged in ICT services with B2B business processes. The research uses the CRISP-DM methodology and is expected to make the sales pipeline business process in CRM more optimal and efficient. By optimizing CRM data with machine learning, companies can focus on sales strategy [12].

## 2. LITERATURE REVIEW

### 2.1 CRM

CRM is a business language, methodology, application software, which describes the customer relationship with the company [18]. Until now there is no definite definition of CRM, and this does not change the general understanding of CRM which ensures the customer cycle can improve company performance [6]. CRM is used to increase revenue, but CRM is not always aimed at increasing revenue [19]. CRM is centered on the customer cycle starting from acquisition, retention, and win strategy [8]. More than just a technology, CRM is a philosophy and a customer cycle strategy. Implementation of CRM applications is among the top investments in technology [20].

#### 2.1.1 B2B CRM

CRM in the B2B sector is critical to marketing and sales success [21]. CRM consolidates all sales activities and distributes them to various company systems [22]. The concept of the B2B CRM Application provides customer-centric management solutions which are the key for companies to maximize customer portfolio [23]. After the customer data is collected, the marketing automation function is executed for the digital marketing management process [24]. And interested customers will enter the sales pipeline, and stakeholders will monitor the sales process [24].

#### 2.1.2 Sales Pipeline B2B CRM

The sales pipeline is part of B2B CRM and is fundamental to sales pipeline management [9]. B2B companies create economic value and experience for customers in the sales process in the sales pipeline [25]. The sales pipeline process starts from collecting data on old and new prospective customers, and some researchers use leads management as an innovative strategy to get new sales opportunities [8], with the stages of opportunity:

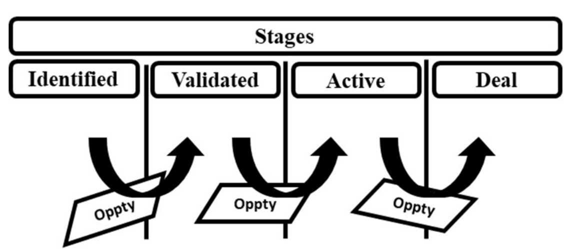


*Figure 1: The Stages of Opportunity*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Interested customers will have prospected as opportunities, and the company determines the stages of opportunities [26]. Identified sales opportunities are categorized if the interested customer is valid [26]. Validated opportunities, the company's process of independently verifying prospective new customers. Furthermore, the prospective customer runs with the category of active opportunities. In the end, the opportunity is determined to be lost or won at the deal opportunity stage [26].

## 2.2 Data Mining

There are some gaps between data analysis requirements and application functionality. To bridge the gap, media for data analysis is needed [27]. With the aim of data can be processed and interpreted visually. Data analytics such as business intelligence, and data mining have clearer paths for better analysis [28].

Data mining has an important role in various human activities to present knowledge [29], with the process of extracting information from transaction data processing [30]. Data mining is also known as Knowledge Discovery in Database (KDD), which is processing large data [31]. There are 2 categories of data mining, namely supervised which is a data mining technique that requires target variables. And Unsupervised doesn't need a target variable [32].

## 2.3 Data Mining - Classification Technique

Big data is an important part of data processing based on statistical processes [33]. The data extraction process can be selected based on the type of data and the variables you want to predict. The classification technique is one of the algorithms used in data mining [34]. Classification is one of the big data processing techniques with machine learning model classifier algorithms [35]. Building a classifier model construction requires a data training phase, each training data set is described by some form of attribute such as nominal data type [36].

Decision tree, Naives Bayes, and k-NN are classification algorithm techniques from several existing techniques. [37], and included in the Top 10 data mining based on a survey [38]. The Decision tree C4.5 algorithm was developed by J. Ross Quinlan in 1993 [39], and is an improvement of the ID3 algorithm. The decision tree model was developed based on the gain entropy information. A higher entropy value indicates a class value preference of the target variable [40]. The Naive Bayes algorithm is called Naïve Beyes because it assumes that all variables contribute to the classification and are correlated with each other [41]. While the k-NN algorithm is the simplest, because of the principle of instance-based learning or lazy learners [41]. Lazy learner algorithms require longer computational time in the training process than decision tree and Naives Bayes [41].

From the survey and comparison analysis of the classification algorithm using UCI data, it is concluded that the decision tree algorithm is more accurate and is an easier algorithm to use [41]

## 2.4 Data Mining Process Model - CRISP-DM

The industrial world requires standard processes to carry out data mining processes that aim to solve business problems [15]. Predictive analysis can be used as an efficiency for companies to maximize time [42]. CRISP-DM, Knowledge Discovery Databases (KDD), and SEMMA are popular process data mining models [15]. with the following comparison:

*Table 1: Comparison of Data Mining Process Models [15]*

| Data Mining Process Models | | | | | |
|---|---|---|---|---|---|
| **KDD (9 Steps)** | | **CRISP-DM (6 Steps)** | | **SEMMA (5 Steps)** | |
| 1 | Developing and Understanding of the Application | 1 | Business Understanding | | - |
| 2 | Creating a Target Data Set | 2 | Business Understanding | 1 | Sample |
| 3 | Data Cleaning and Pre-processing | | | 2 | Explore |
| 4 | Data Transformation | 3 | Data Preparation | 3 | Modify |
| 5 | Choosing the suitable Data Mining Task | 4 | Modeling | 4 | Model |
| 6 | Choosing the suitable Data Modeling Model Mining Algorithm | | | | |
| 7 | Employing Data Mining Algorithm | | | | |
| 8 | Interpreting Mined Patterns | 5 | Evaluation | 5 | Assessm ent |
| 9 | Using Discovered Knowledge | 6 | Deployment | | - |

Journal of Theoretical and Applied Information Technology
31st May 2022. Vol.100. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From the results of the comparative analysis of the data mining model process, CRISP-DM is the most complete model [15]. CRISP-DM has a cycle that is divided into 6 stages [15] [32] :

a. Business understanding, the phase of determining the goals of the data mining project to be carried out.
b. Data understanding is the phase that aligns data mining goals with analysis and evaluation.
c. Data preparation, data collection stage relevant to project objectives.
d. Modeling is the phase of determining the parameters to be used and selecting the algorithm.
e. Evaluate, evaluate, define and make decisions on the results of the data mining process.
f. Deployment, implementation phase and use the resulting model.

## 3. RESEARCH METHODOLOGY

The following is the research methodology of this study with the adoption of the CRISP-DM methodology:

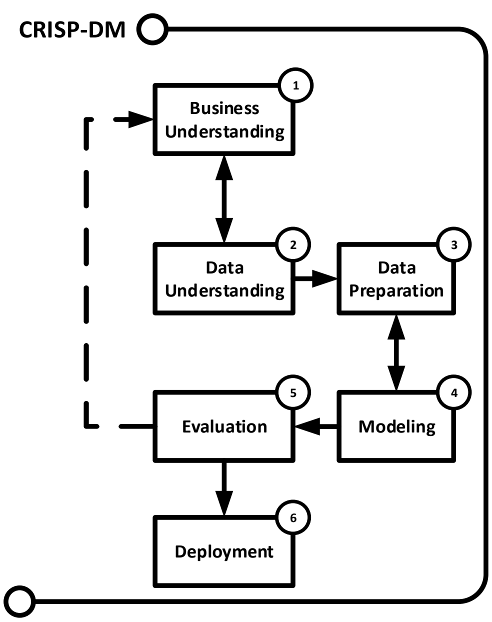


*Figure 2: CRISP-DM Cycle Model*

### 3.1. Business understanding

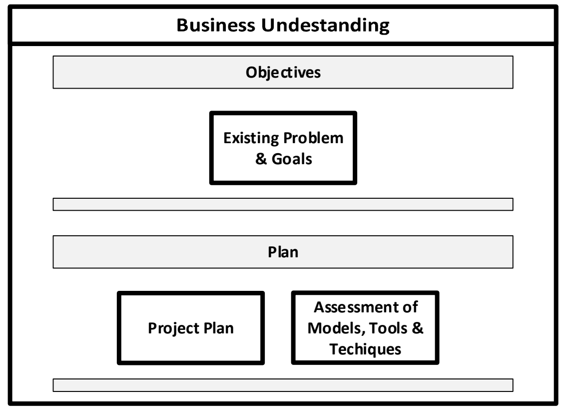


*Figure 3: Phase 1 - Business Understanding*

In the initial phase, an analysis of existing business processes is carried out, which aims to understand the problems and why a data mining project is needed to be related to the determination of lost or won opportunities. From the analysis process, it is hoped that there will be a solution to improve the sales business process. Next, project initiation begins with preparing a project plan and conducting an assessment related to the models, techniques, and tools that will be used.

### 3.2. Data understanding

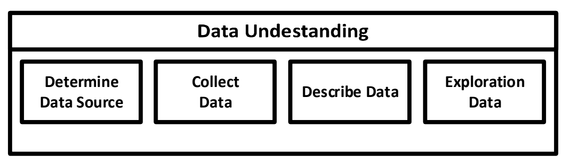


*Figure 4: Phase 2 - Data Understanding*

The source data is taken from the B2B CRM application, because the sales pipeline, customer profile data, and sales history are included in the process of the B2B CRM application. The data is then given descriptive information. The data variables are then explored based on the description information to determine the relevance of the data mining project to be carried out.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### 3.3. Data preparation

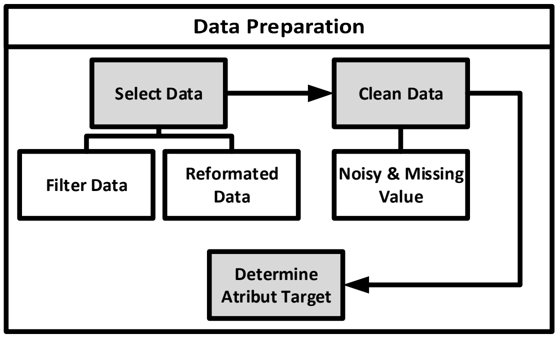


*Figure 5: Phase 3 - Data Preparation*

Furthermore, the data filter process is carried out which aims to filter the data that will be used. The filtered data is carried out by a cleansing process so that anomaly data can be deleted or data imputation is carried out. The last preparation is the process of setting the format and role variables.

### 3.4. Modeling

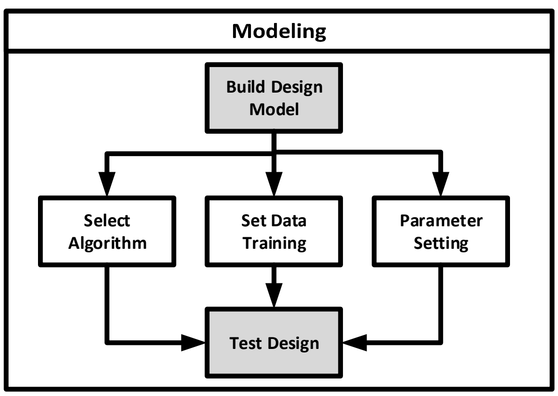


*Figure 6: Phase 4 - Modeling*

The modeling process is carried out using machine learning tools. Before the design test, the process of selecting a supervised learning classification algorithm was carried out according to the data type and target variable. In this study uses a decision tree algorithm. The dataset is divided into two, namely training data and testing data. In addition, to test design, the K-Fold Cross Validation test is carried out to assess the performance of the selected algorithm by dividing the sample randomly and classy data as much as K-Fold.

### 3.5. Evaluation

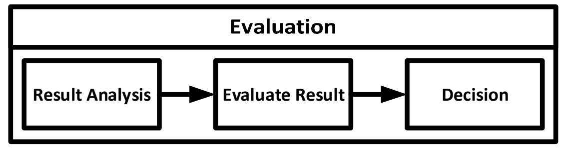


*Figure 7: Phase 5 - Modeling*

The process of evaluating the design test is done by analyzing the percentage of accuracy with the confusion matrix table and AUC Curve. The results of the analysis are evaluated whether they are by the project objectives.

*Table 2: Confusion Matrix*

| | | True Values | |
|---|---|---|---|
| | | **TRUE** | **FALSE** |
| **Prediction** | **TRUE** | **True Positive (TP)** (Correct Result) | **False Positive (FP)** ( Unexpected Result) |
| | **FALSE** | **False Negative (FN)** (Missing Result) | **True Negative (TN)** (Correct Absence of Result) |

$$\textbf{Accuracy:} \quad \frac{TP + TN}{TP + TN + FP + FN}$$

ROC curve is a technique for visualizing and classifying results based on performance, with the basic guidelines described as follows [43] :

*Table 3: ROC Curve Explanation [43]*

| Scale | Explanation Classification |
|---|---|
| 0,90 – 1,00 | Excellent |
| 0,80 – 0,90 | Good |
| 0,70 – 0,80 | Fair |
| 0,60 – 0,70 | Poor |
| 0,50 – 0,60 | Failure |

### 3.6. Deployment

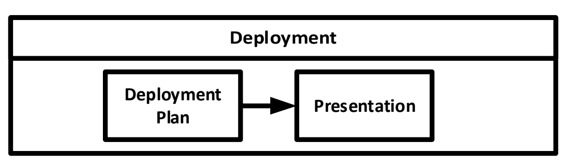


*Figure 8: Phase 6 - Modeling*

The deployment plan was presented explaining how machine learning can be implemented in current B2B CRM applications by prototyping.

## 4. RESULT & DISCUSSION

### 4.1 Business Understanding

The object of research is in one of the B2B companies engaged in the ICT sector and the data is taken from one of the private companies in the ICT sector. And the data is taken form the internal CRM application in the company. The following are the stages of the sales pipeline in the company's B2B CRM application:

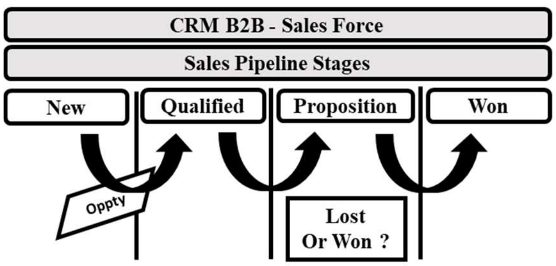


*Figure 9: Sales Pipeline Current Stages*

Stage explanation:

1. Stage New is a valid Opportunity from potential new customers and existing customers who are interested in the products and solutions offered by the company.
2. Stage Qualified is the stage of further prospects of customer interest. At this stage, the company makes commercial offer proposals and technical proposals for solutions related to the products and services offered.
3. Proposition Stage is the stage of determining the lost or won the opportunity, whether the opportunity is won or vice versa. The determination process comes from a direct negotiation process or tender competition with competitors.
4. Stage Won is a sales opportunity data that has been won and becomes a work project.

From the description of the staging in the company, lost or won is in the Proposition stages. The current problem arises that the time for preparing project resources is very short. The second problem is that the opportunity with a low probability of winning is not monitored from the start and the possibility of losing the opportunity is very large. From these problems, predictive data processing is needed to monitor and reduce the risk of current problems.

### 4.2 Business Understanding

The data source taken comes from several sources of B2B CRM application modules, such as the sales force module, project module, and customer management module. The data source is taken by exporting excel and CSV format data from each module.

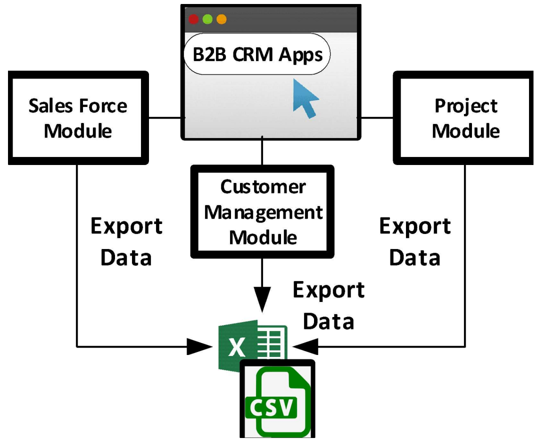


*Figure 10: Export data from B2B CRM Applications*

The initial field attributes taken from the data source are 15 attributes. The number of rows of initial data exported was 654 rows of data from the company's B2B sales transactions from 2020 to 2021.

*Table 4: Description attribute variable*

| No | Attribute Name (Source Module) | Attribute Description |
|---|---|---|
| 1 | Opportunity (SalesForce) | Explanation of the name of the opportunity that is being prospected |
| 2 | Customer (Customer Management) | A valid prospect customer who opens a sales opportunity |
| 3 | Segment (Customer Management) | Customer categories based on company status |
| 4 | Sub Segment (Customer Management) | Business segment customers |
| 5 | Amount (SalesForce) | The estimated value of the projected gross profit of the opportunity |
| 6 | Sales Team (SalesForce) | Sales team responsible for sales opportunity |
| 7 | Sales Subteam (SalesForce) | Sales grouping by business segment |
| 8 | Product (SalesForce) | The product or service solution offered |
| 9 | PreSales/Name (SalesForce) | Technical personnel who helps explain solutions related to products and services |

| No | Attribute Name (Source Module) | Attribute Description |
|---|---|---|
| 10 | Salesperson/Name (SalesForce) | Commercial personnel who processes prospects, engagements, and price offers to customers |
| 11 | Stage Status (SalesForce) | Sales opportunity status, lost or won |
| 12 | Competitor (SalesForce) | Information on competitors on opportunities that are being prospected |
| 13 | Commit (SalesForce) | Information on the sales person's commitment to the opportunity |
| 14 | Recurring (Project) | Project Data derived from customer order schedule |
| 15 | Aging Sales (SalesForce) | Aging information based on the total length of the sales process |

### 4.3 Data Preparation

The filtering process is started by filtering the relevant data. The recurring data filter = yes is not used because the data is not an opportunity that goes through the sales pipeline stages. Attributes whose row data is empty will not be used, such as competitor data. From the results of the data preparation process, the variable attributes used are 10 attributes, and the B2B transaction data row becomes 518 Row data.

*Table 5: Attribute variables used*

| No | Attribute Name (Source Module) | Role Type |
|---|---|---|
| 1 | Opportunity (Sales Force) | Id |
| 2 | Customer (Customer Management) | - |
| 3 | Segment (Customer Management) | - |
| 4 | Sub Segment (Customer Management) | - |
| 5 | Amount (SalesForce) | - |
| 6 | Sales Team (SalesForce) | - |
| 7 | Product (SalesForce) | - |
| 8 | PreSales/Name (SalesForce) | - |
| 9 | Salesperson/Name (SalesForce) | - |
| 10 | Stage Status (SalesForce) | Label Target |

*Table 6: Dataset Label Target Fraction*

| Label Target | Count | Fraction |
|---|---|---|
| Lost | 291 | 57% |
| Won | 227 | 43% |
| **Total** | **518** | **100%** |

### 4.4 Modeling

The modeling process is carried out before the test design is carried out. Modeling and testing designs are processed using machine learning tools. The steps of the modeling process are:
The chosen algorithm is a decision tree with gain ratio criteria.

1. Algorithm performance tested with cross-validation K=10 Fold
2. Divide the dataset into 2, namely training data and testing data (70:30)
3. The distribution of training data and testing data uses a stratified sampling configuration, which is to build a random subset and ensure that the distribution of classes in the subset is the same as throughout the dataset.

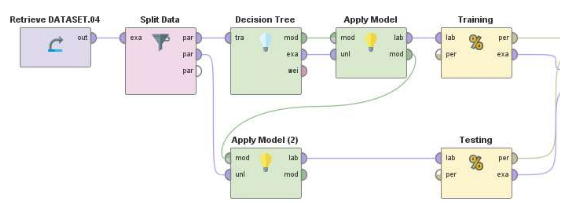


*Figure 11: Test design modeling using rapidminer machine learning tools*

### 4.5 Evaluation

From the modeling results, algorithm performance tests and design tests were carried out using rapidminer machine learning tools, and the accuracy results from the confusion matrix and performance based on the AUC Curve were as follows:

Journal of Theoretical and Applied Information Technology
31st May 2022. Vol.100. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**1. Test results cross validation10-fold:**

*Table 7: Analysis of confusion matrix cross-validation 10-fold*

| | true Lost | true Won | class precision |
|---|---|---|---|
| **pred. Lost** | 242 | 94 | 72.02% |
| **pred. Won** | 49 | 133 | 73.08% |
| **class recall** | 83.16% | 58.59% | |

$$\textbf{Accuracy:} \quad \frac{242 + 133}{242 + 133 + 94 + 49} = \textbf{72.39 \%}$$

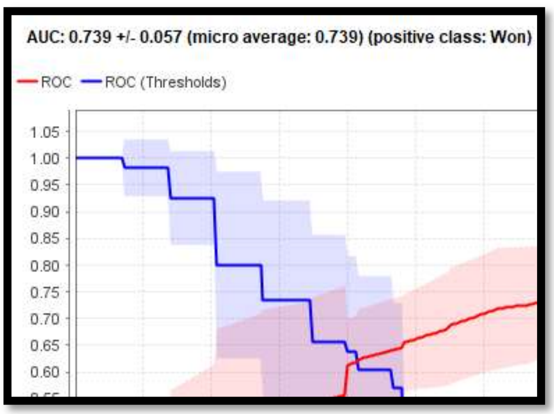


*Figure 12: AUC curve cross-validation 10-fold*

AUC cross-validation 10-fold = 0.739 (Fair Classification)

**2. The results of the training data test:**

*Table 8: Analysis of confusion matrix data training*

| | true Lost | true Won | class precision |
|---|---|---|---|
| **pred. Lost** | 192 | 44 | 81.36% |
| **pred. Won** | 12 | 115 | 90.55% |
| **class recall** | 94.12% | 72.33% | |

$$\textbf{Accuracy :} \quad \frac{192 + 115}{192 + 115 + 44 + 12} = \textbf{84.57 \%}$$

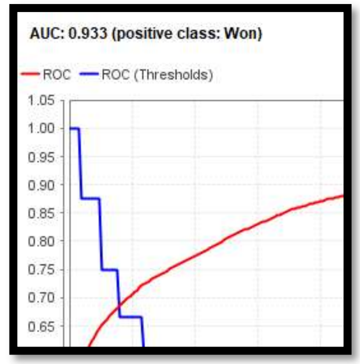


*Figure 13: AUC Curve data training*

AUC data training = 0.933 (Excellent Classification)

**3. Test results of data testing:**

*Table 9: Analysis of confusion matrix data testing*

| | true Lost | true Won | class precision |
|---|---|---|---|
| **pred. Lost** | 71 | 36 | 66.36% |
| **pred. Won** | 16 | 32 | 66.67% |
| **class recall** | 81.61% | 47.06% | |

$$\textbf{Accuracy :} \quad \frac{71 + 32}{71 + 32 + 36 + 16} = \textbf{66.45 \%}$$

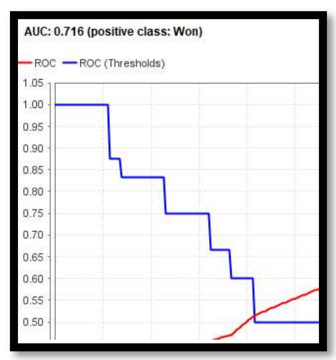


*Figure 14: AUC Curve data testing*

AUC data testing = 0.716 (Fair Classification)

From the evaluation results, it is found that the accuracy of the decision tree algorithm gain ratio is 72.39%, while the accuracy of the testing data is 66.45%. Performance according to the calculation of the AUC curve is in the fair classification for cross-validation testing and data testing. Based on the results of the performance evaluation and accuracy, the design modeling is acceptable and can be implemented in the sales pipeline process in the B2B CRM application.

### 4.6 Deployment

The deployment plan is represented by an architectural description of the integration between B2B CRM applications and machine learning tools, then a prototyping description of the input data opportunity form in the sales force module is made, and business analysis of the sales staging process from integration with machine learning in the sales pipeline.

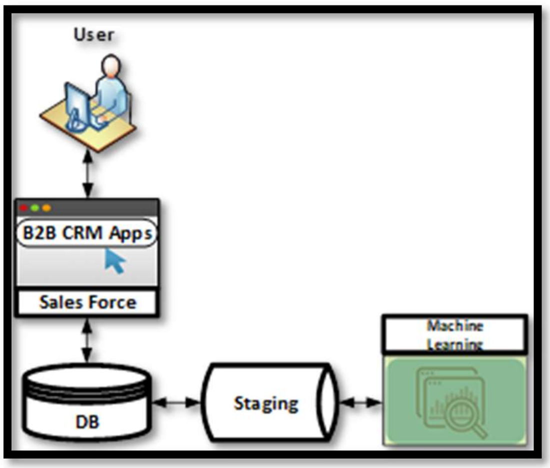


*Figure 15: High-level architecture*

The description of the high-level architecture explains, When the user inputs the opportunity data, the data will be sent to staging. Then machine learning will carry out the process of getting data to process the prediction analysis of lost or won opportunities in the form of percentage probability. The prediction processing data will be sent to staging, and the B2B CRM application will retrieve the data to be visualized in the data opportunity in the sales force module.

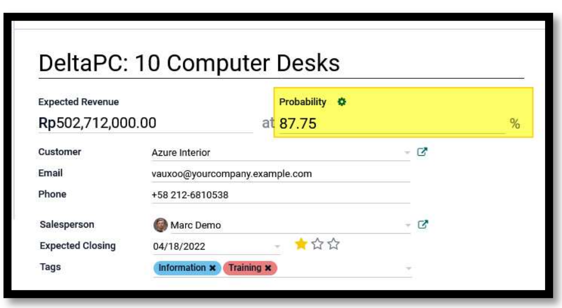


*Figure 16: Opportunity data form prototype*

The probability visualization will appear on the opportunity detail data input form. The probability picture can be used as a user reference to develop a sales strategy related to the probability value that appears in the opportunity data.

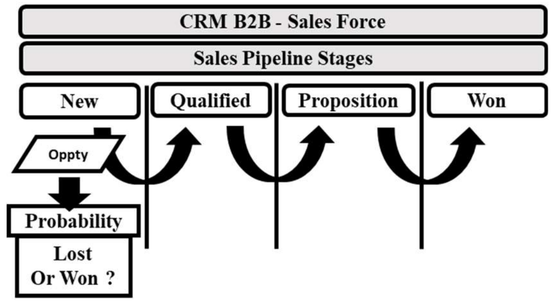


*Figure 17: Analysis sales pipeline staging after implementation machine learning*

From the results of probability machine learning processing, if the prediction or probability lost or won opportunity can be identified when the opportunity data is in the early stages of the sales pipeline.

## 5. CONCLUSION

From the results of the presentation of the results and discussion, it is concluded that with the help of machine learning processes in the B2B CRM application. The business process at the sales pipeline stage of B2B CRM can be improved with the probability lost or won can be estimated when the opportunity is first verified. So that companies can minimize the risk of lost opportunities. This is because the company will monitor and control opportunities that have a low probability of winning. Another thing, the company has projections related to planning for future resource requirements that will run the project when the opportunity is won.

The B2B CRM data processing process with machine learning can be implemented in all companies with a B2B sales model business. Because the B2B sales model business has several stages of administration and negotiation for a long time before the customer buys the product or service offered.

## REFERENCES:

[1] A. K. Rathore, S. Shioramwar, and P. V. Ilavarasan, "Social customer relationship management as predictor of customer relationship performance: An empirical study of B2B and B2C companies," 2016, doi: 10.1007/978-3-319-45234-0_16.

[2] A. Ahani, N. Z. A. Rahim, and M. Nilashi, "Forecasting social CRM adoption in SMEs: A combined SEM-neural network method," *Comput. Human Behav.*, 2017, doi: 10.1016/j.chb.2017.05.032.

[3] W. T. Venturini and S. G. Benito, "CRM software success: A proposed performance measurement scale," *J. Knowl. Manag.*, 2015, doi: 10.1108/JKM-10-2014-0401.

[4] Z. Soltani and N. Jafari, "Computers in Human Behavior Customer relationship management mechanisms : A systematic review of the state of the art literature and recommendations for future research," *Comput. Human Behav.*, vol. 61, pp. 667–688, 2016, doi: 10.1016/j.chb.2016.03.008.

[5] R. S. Hassan, A. Nawaz, M. N. Lashari, and F. Zafar, "Effect of Customer Relationship Management on Customer Satisfaction," *Procedia Econ. Financ.*, 2015, doi: 10.1016/s2212-5671(15)00513-4.

[6] V. Eitle and P. Buxmann, "Business analytics for sales pipeline management in the software industry: A machine learning perspective," 2019, doi: 10.24251/hicss.2019.125.

[7] J. Chen, "Business-to-Business (B2B) Definition," *Investopdia*, 2020. .

[8] P. Espadinha-Cruz, A. Fernandes, and A. Grilo, "Lead management optimization using data mining: A case in the telecommunications sector," *Comput. Ind. Eng.*, 2021, doi: 10.1016/j.cie.2021.107122.

[9] J. Yan, M. Gong, C. Sun, J. Huang, and S. M. Chu, "Sales pipeline win propensity prediction: A regression approach," 2015, doi: 10.1109/INM.2015.7140393.

[10] M. J. Hamid Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *Int. J. Adv. Comput. Sci. Appl.*, 2018, doi: 10.14569/IJACSA.2018.090630.

[11] N. Legowo and K. Hwibawa, "Data mining as tools to improve marketing campaign," *Int. J. Adv. Trends Comput. Sci. Eng.*, 2020, doi: 10.30534/ijatcse/2020/57952020.

[12] F. Guo and H. Qin, "Data Mining Techniques for Customer Relationship Management," 2017, doi: 10.1088/1742-6596/910/1/012021.

[13] D. Rotovei and V. Negru, "Improving lost/won classification in CRM systems using sentiment analysis," 2018, doi: 10.1109/SYNASC.2017.00038.

[14] W. Y. Ayele, "Adapting CRISP-DM for idea mining a data mining process for generating ideas using a textual dataset," *Int. J. Adv. Comput. Sci. Appl.*, 2020, doi: 10.14569/IJACSA.2020.0110603.

[15] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models ( KDD , CRISP-DM and SEMMA )," *Int. J. Innov. Sci. Res.*, 2014.

[16] A. Khumaidi, "DATA MINING FOR PREDICTING THE AMOUNT OF COFFEE PRODUCTION USING CRISP-DM METHOD," *J. Techno Nusa Mandiri*, 2020, doi: 10.33480/techno.v17i1.1240.

[17] G. Soliman, A. El-Nabawy, A. Misbah, and S. Eldawlatly, "Predicting all star player in the national basketball association using random forest," 2018, doi: 10.1109/IntelliSys.2017.8324371.

[18] W. John, "Modelling CRM applications and its impact on relationship outcomes: A B2B relationship perspective," *Int. J. Electron. Cust. Relatsh. Manag.*, 2018, doi: 10.1504/IJECRM.2018.096240.

[19] S. J. Kinnett, *How to Win at CRM*. 2017.

[20] R. M. Peterson, M. Rodriguez, and V. Krishnan, "Crm and Sales Pipeline Management : Empirical Results for Managing Opportunities," *Mark. Manag. J.*, 2011.

[21] Y. E. Zeng, H. Joseph Wen, and D. C. Yen, "Customer relationship management (CRM) in business-to-business (B2B) e-commerce," *Inf. Manag. Comput. Secur.*, 2003, doi: 10.1108/09685220310463722.

[22] J. P. . Laudon and K. C. . Laudon, "MANAGEMENT INFORMATION SYSTEMS : Managing the Digital Firm Laudon and Jane P . Laudon ( Book Review )," *Int. J.*, 2007.

[23] U. Zeynep Ata and A. Toker, "The effect of customer relationship management adoption in business-to-business markets," *J. Bus. Ind. Mark.*, 2012, doi: 10.1108/08858621211251497.

[24] A. Cruz and A. Vasconcelos, "A reference application architecture for the CRM domain: The Portuguese citizen space case study," *Int. J. Enterp. Inf. Syst.*, 2015, doi: 10.4018/IJEIS.2015040102.

[25] B. J. Pine, "How B2B companies create economic value by designing experiences and transformations for their customers," *Strateg. Leadersh.*, 2015, doi: 10.1108/SL-03-2015-0018.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195


[26] A. Megahed, P. Yin, and H. R. M. Nezhad, "An optimization approach to services sales forecasting in a multi-staged sales pipeline," 2016, doi: 10.1109/SCC.2016.98.

[27] S. Wikner, "The Next Generation CRM Tools: Bridging the Gaps between Sales Needs and CRM Tools Architecture," in *Organizing Marketing and Sales*, 2018.

[28] R. Dale Wilson, "Developing new business strategies in B2B markets by combining CRM concepts and online databases," *Compet. Rev.*, 2006, doi: 10.1108/cr.2006.16.1.38.

[29] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *Int. J. Inf. Technol.*, 2020, doi: 10.1007/s41870-020-00427-7.

[30] W. M. P. Van Der Aalst and M. Song, "Mining Social Networks : Uncovering Interaction," *Science (80-. ).*, 2004.

[31] M. Bramer, "Principles of Data Mining, ser. Undergraduate Topics in Computer Science," *Vasa*, 2007.

[32] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. 2014.

[33] B. Çiğşar and D. Ünal, "Comparison of Data Mining Classification Algorithms Determining the Default Risk," *Sci. Program.*, 2019, doi: 10.1155/2019/8706505.

[34] M. Saouabi and A. Ezzati, "Data mining classification algorithms," *Int. J. Math. Comput. Sci.*, 2020.

[35] K. I. Taher, A. M. Abdulazeez, and D. A. Zebari, "Data Mining Classification Algorithms for Analyzing Soil Data," *Asian J. Res. Comput. Sci.*, 2021, doi: 10.9734/ajrcos/2021/v8i230196.

[36] A. Urso, A. Fiannaca, M. La Rosa, V. Ravì, and R. Rizzo, "Data mining: Classification and prediction," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 2018.

[37] M. R. K. Hofmann, "Data Mining Use Cases and Business Analytics Applications," 2016.

[38] W. C. Lin, S. W. Ke, and C. F. Tsai, "Top 10 data mining techniques in business applications: a brief survey," *Kybernetes*, vol. 46, no. 7, pp. 1158–1170, 2017, doi: 10.1108/K-10-2016-0302.

[39] B. HSSINA, A. MERBOUHA, H. EZZIKOURI, and M. ERRITALI, "A comparative study of decision tree ID3 and C4.5," *Int. J. Adv. Comput. Sci. Appl.*, 2014, doi: 10.14569/specialissue.2014.040203.

[40] W. F. W. Yaacob, S. A. M. Nasir, W. F. W. Yaacob, and N. M. Sobri, "Supervised data mining approach for predicting student performance," *Indones. J. Electr. Eng. Comput. Sci.*, 2019, doi: 10.11591/ijeecs.v16.i3.pp1584-1592.

[41] H. P. C. Sayali D. Jadhav, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *Int. J. Sci. Res.*, 2016, doi: 10.21275/v5i1.nov153131.

[42] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," 2019, doi: 10.1016/j.procir.2019.02.106.

[43] F. Gorunescu, "Data mining: Concepts, models and techniques," *Intell. Syst. Ref. Libr.*, 2011, doi: 10.1007/978-3-642-19721-5.