Bussines Intelligent in Telemarketing Using SVM

Putu Agung Ananta Wijaya[1], Komang Budiarta[2] , and Made Sudarma[3]

[1][2] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University

[3] Department of Electrical and Computer Engineering, Udayana University

Email: ananta_wijaya@student.unud.ac.id

Abstract— **Direct marketing provides an advantage in approaching consumers. Communication that happens allows us more closely, able to change the behavior and know the needs required by consumers accurately. But this technique has a lack of time. It takes a long time to convince consumers to buy the products offered. Bussines intelligent with data mining approach to consumer data is required. This process will analyze the potential possessed by a consumer. At the stage of the DSS used SVM method to predict whether consumers will buy products that have been offered. Bussines intelligent built proven able to predict consumers who have the potential to buy products. Tests show the greatest prediction accuracy rate is 89.5% with a combination of data traning of 70% of the dataset.**

Index Term—Data Mining, Linier SVM, Holdout

I. INTRODUCTION

Marketing is a series of activities to plan, promote, and distribute goods or services to consumers in order to meet the needs of consumers.Marketing is about identifying and meeting human and social needs[1]. One of the shortest good definition of marketing is meeting needs profitably. To convey the value of the product to the customer is required communication. Marketing communications are the means by which the company in an effort to inform, persuade, and remind consumers directly or indirectly about the product or brand they sell [1]. Communication method on marketing is divided into two, indirect marketing and direct marketing.

Indirect marketing is a strategy to promote a product or service intended to touch the mind and feelings of consumers indirectly. Indirect sales can be found in the form of advertising, corporate social responsibility, and interactive marketing via the internet. Direct Marketing is an interactive marketing system that utilizes one or more advertising media to generate measurable responses and or transactions in any location [2]. One form of direct marketing is telemarketing [3].

Direct marketing provides an advantage in approaching consumers. Communication that happens allows us more closely, able to change the behavior and know the needs required by consumers accurately. With direct marketing the company will more quickly know the results of marketing done. But this technique has a lack of time. It takes a long time to convince consumers to buy the products offered.

Business intelligence with data mining approach to consumer data is needed. Business Intelligence is comprises the set of strategies, processes, applications, data, technologies and technical architectures which are used by enterprises to support the collection, data analysis, presentation and dissemination of business information[4]. Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner[5].

The system to be built has a purpose to analyze the potential possessed by a consumer. In the DSS stage used SVM method (Support Vector Machines) to predict whether consumers will buy the products that have been offered. SVM is a pattern recognition method that is classified as supervised. In a two-class learning task, the aim of SVM is to fi nd the best classi fi cation function to distinguish between members of the two classes in the training data [6]. This method offers great resilience and accuracy among other popular methods [7].

With the construction of this system is expected to make a decision about whether or not the marketing process can be taken faster. And when it comes to consumers who have the potential to buy products then it can be compiled more quickly the appropriate marketing strategy.

II. RELATED RESEARCH

A. Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination

In this research will analyze telemarketing dataset from UCI Repository. The method of classification used is the Naive Bayes algorithm. This method has a purpose to predict the decision of Telemarketing clients. The method of classification will be combined with the feature selection method. The feature selection method used is Backward Elimination. The test results show that the accuracy of Naive Bayes 89.08%, after the selection of features using Backward Elimination obtained a higher accuracy of 90.69%, by looking at the accuracy of the algorithm Naive Bayes-based Backward Elimination increase accuracy to predict Decision of Telemarketing clients [8].

B. Optimasi Algoritma Support Vector Machine (SVM) Menggunakan Adaboost Untuk Penilaian Risiko Kredit

Support Vector Machine (SVM) is an algorithm proposed by many researchers in the field of data mining credit risk. But the difficulty of determining the parameters of an ideal is the problem of research in improving the accuracy of SVM. In this

study, the use of AdaBoost is proposed to improve the accuracy of SVM with the selection of the kernel, the value of the parameter C, and the appropriate iteration. The results showed that the use of Boosting (AdaBoost-SVM) has better accuracy than Bagging optimization model (Bagging-SVM) and SVM with conventional C's election. Using AdaBoost able to correct errors on the base SVM classifiers. And SVM without optimization into the model with the lowest accuracy followed by Bagging-SVM[9].

C. Penerapan Algoritma Genetika untuk Optimasi Parameter pada Support Vector Machine untuk Meningkatkan Prediksi Pemasaran Langsung

In this study, researchers used SVM methods for predicting direct marketing results. In addition, genetic algorithms are used for parameter optimization. Several experiments were conducted to obtain optimal accuracy. The results showed that experiments using support vector machine and genetic algorithm used to optimize parameters C, γ and ε with three types of kernel. Kernel first kernel dot type with an accuracy of 85.59%, AUC of 0.911 which both types of radial kernel with an accuracy of 98.89%, AUC of 0.981 and the third with Polynomial kernel type with an accuracy of 98.67% and AUC of 0.938. The experimental results show that the test data set using the application of genetic algorithm to the support vector machine shows more accurate results for direct marketing predictions [10].

III. LITERATURE REVIEW

A. Data Mining

Data Mining (DM) is the process of analyzing data from various perspectives and summarize it into useful information, where the information can be used to increase revenue, cut expenses or both [11]. The main tasks that can be done by data mining are (1) Description, (2) Estimation, (3) Prediction, (4) Classification, (5) Clustering and (6) Association [12]. Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data, or KDD, while others view data mining as merely an essential step in the process of knowledge discovery. The knowledge discovery process is shown in Figure 1 as an iterative sequence of the following steps[13]:

1. Cleaning the data (to remove the data are inconsistent and noise)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations)
5. Data mining (an essential process where intelligent methods are applied to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interestingness measures)

7. Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users)
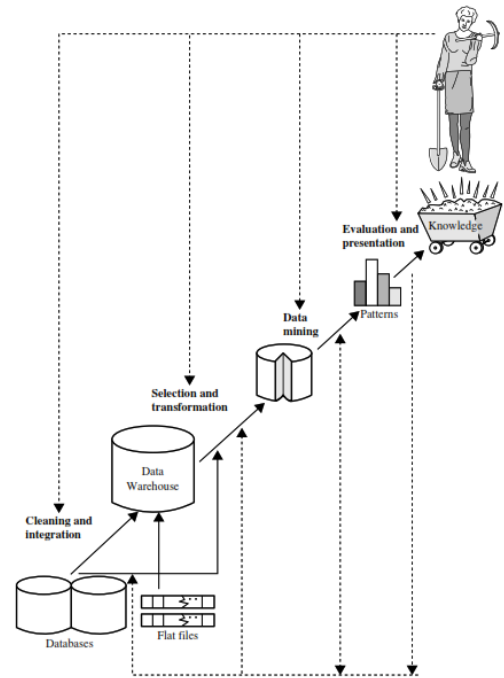

Fig. 1. Steps of Data Mining

B. SVM (Support Vector Machine)

Pattern Recognition is one of the fields in computer science, which maps data into certain predefined concepts. This particular concept is called a class or category. The application of pattern recognition is very broad, including recognizing the sound in the security system, reading the letters in the OCR, classifying the disease automatically based on the diagnosis of the patient's medical condition and so on. Various methods are known in recognition patterns, such as linear

SVM is a learning machine method that works on the principle of Structural Risk Minimization (SRM) in order to find the best hyperplane that separates the two classes in the input space. Unlike the neural network strategy that seeks hyperplane separators between classes, SVM attempts to find the best hyperplane in input space [6].
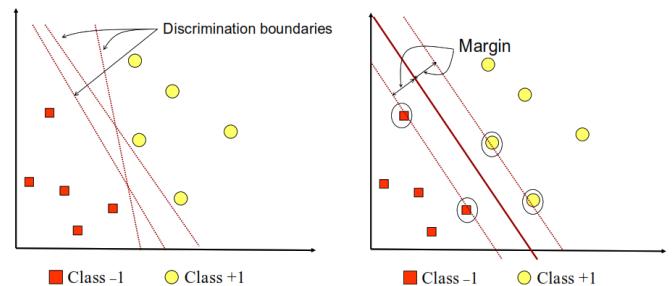

Fig. 2. SVM's illustration

The concept of SVM can be explained simply as a search for the best hyperplane that serves as a separator of two classes in the input space. Fig 2 shows some patterns that are members of two classes: +1 and -1. Patterns belonging to class -1 are symbolized in red (box), while the pattern in class +1, symbolized by yellow (circle). Classification problems can be translated by finding the line (hyperplane) that separates the two groups. Various alternate lines of discrimination (discrimination boundaries).

The best separator hyperplane between the two classes can be found by measuring the hyperplane margin. And look for the maximum point. Margin is the distance between the hyperplane and the nearest pattern of each class. The closest pattern is called a support vector. The solid line in Figure 1-b shows the best hyperplane, which is located right in the middle of both classes, while the red and yellow dots that are in the black circle are the support vector. The effort to locate the hyperplane is at the core of the learning process in SVM.

Available data is denoted as $\vec{x_i} \in \Re^d$ while each label is denoted as $y_i \in \{-1, +1\}$ for i = 1, 2, ..., l where l is the number of data. It is assumed that both classes -1 and +1 can be perfectly separated by the dimensionless hyperlplane d, which is defined

$$\vec{w} . \vec{x} + b = 0 \qquad (1)$$

Patterns $\vec{x_i}$ that belong to class -1 can be defined as a pattern that satisfies inequality.

$$\vec{w} . \vec{x_i} + b \leq -1 \qquad (2)$$

Patterns $\vec{x_i}$ that belong to class +1 can be defined as

$$\vec{w} . \vec{x_i} + b \geq +1 \qquad (3)$$

The greatest margin can be found by maximizing the distance value between the hyperplane and its nearest point ie $1/{\|\vec{w}\|}$.

C. Dataset

In this study used dataset from University of California. The dataset used is * .CSV type with bank.csv name. The number of datasets is 4521 lines with 17 attributes where the last attribute is a class of data. This data is downloaded at: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing The data is related to the direct marketing campaigns of a Portuguese banking institution. The marketing campaigns are based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed [14].

IV. DISCUSSION

To get the desired results each step of data mining is done carefully. Here are the steps of this research:
1. Cleaning the data
   The data cleaning process is done by deleting rows that have attribute value "null"
2. Data Integration

This process is not done because the data used is bank.csv dataset from University of California.
3. Data Selection
   All the columns (variables) contained in the dataset are used in this study. By using Holdout method separated data traning with data testing.
4. Data Transformation
   Data used in the format from * .csv to * .mat. If represented in matrix data form the size of matrix [4521 * 17]. Then done normalization of data by using normalization min max.
5. Data Mining
   Implement SVM on the dataset. The strategy of this study is to change the value of holdout to see the level of accuracy obtained.
6. Pattern Evaluation
   Mapping accuracy results into confusion matrix.
7. Knowledge presentation
   Mapping SVM results into line graphs and parallel cordinates graphs.

The results of this classification then analyze using based on confusion matrix to measure the performance of the system is based on a confusion table. Confusion matrix are the tables of result classification.

TABLE I.
CONFUSION MATRIX

F_{ij}		Class prediction results (J)	
		Class = -1 (no / not subscribed)	Class = +1 (yes / subscribed)
The original class(i)	Class = -1 (no / not subscribed)	F11	F12
	Class = +1 (yes / subscribed)	F21	F22

Quantity confusion matrix can be summarized into two values, namely accuracy and error rate. By knowing the amount of data that are classified correctly, we can know the accuracy of the results of the predictions and knowing the amount of data that are classified wrong, we can know the error rate from the prediction that done. This two quantities are used as classification performance metrics. To calculate the accuracy used formula 4.

$$a = \frac{amount\ of\ data\ that\ is\ predicted\ correct}{amount\ data\ testing}$$
$$= \frac{F11 + F22}{F11 + F12 + F21 + F22} \qquad (4)$$

a : accuracy
F11 : data with 'no' class that is predicted correct
F12 : data with 'no' class that is predictied wrong
F21 : data with 'yes' class that is predicted wrong
F22 : data with 'yes' class that is predicted correct
To calculate the rate of eror used formula 5.

$$e = \frac{\text{The amount of data that is predicted wrong}}{\text{amount data testing}} \quad (5)$$

$$= \frac{F12 + F21}{F11 + F12 + F21 + F22}$$

e = rate of error

In the first test, the result of prediction is 88,9% with comparison of data traning and data testing that is 10% and 90%. With an error rate of 11.1%. The following in Table II is the result of the first test.

TABLE II.
FIRST TESTING'S CONFUSION MATRIX

F_{ij}		Class prediction results (J)	
		Class = -1 (no / not subscribed)	Class = +1 (yes / subscribed)
The original class(i)	Class = -1 (no / not subscribed)	393	7
	Class = +1 (yes / subscribed)	43	9

In the second test at Holdout validation stage with comparison of data traning and data testing that is 3843 (85%) and 678 (15%). From this test the predicted results are similar to the first test.

TABLE III.
SECOND TESTING'S CONFUSION MATRIX

F_{ij}		Class prediction results (J)	
		Class = -1 (no / not subscribed)	Class = +1 (yes / subscribed)
The original class(i)	Class = -1 (no / not subscribed)	594	6
	Class = +1 (yes / subscribed)	67	11

The third test is done by dividing the dataset by 80% into data traning and the remaining 20% used as data testing. The predicted result is 89.6% and the error rate of 10.4% is shown in table IV.

TABLE IV.
THIRD TESTING'S CONFUSION MATRIX

F_{ij}		Class prediction results (J)	
		Class = -1 (no / not subscribed)	Class = +1 (yes / subscribed)
The original	Class = -1 (no / not subscribed)	787	13
class(i)	Class = +1 (yes / subscribed)	81	23

The fourth test on Holdout validation used data traning of 75% data traning and 25% data testing of the dataset. From this test, the result of prediction is 88,9% presented in table V.

TABLE V.
FOURTH TESTING'S CONFUSION MATRIX

F_{ij}		Class prediction results (J)	
		Class = -1 (no / not subscribed)	Class = +1 (yes / subscribed)
The original class(i)	Class = -1 (no / not subscribed)	991	9
	Class = +1 (yes / subscribed)	116	14

In the fifth test at Holdout validation stage with comparison of data traning and data testing that is 3391 (70%) and 1130 (30%). From this test, the result is predicted 89.5% with error rate of 11.5% presented in table VI.

TABLE VI.
FIFTH TESTING'S CONFUSION MATRIX

F_{ij}		Class prediction results (J)	
		Class = -1 (no / not subscribed)	Class = +1 (yes / subscribed)
The original class(i)	Class = -1 (no / not subscribed)	986	14
	Class = +1 (yes / subscribed)	106	25

Of the five testing strategies that have been done, the highest predictive value with holdout validation with comparison of data traning and data testing is 3617 (80%) and 904 (20%). If all the tests described would be like figure 3.
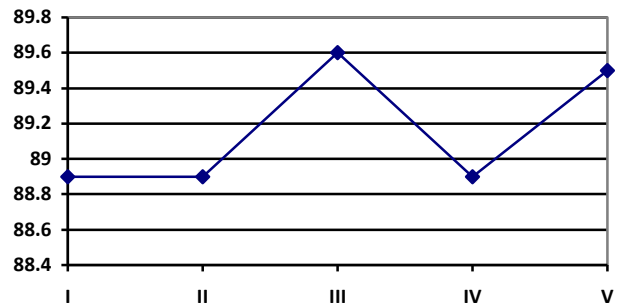

Fig. 3. Prediction Rate Graph

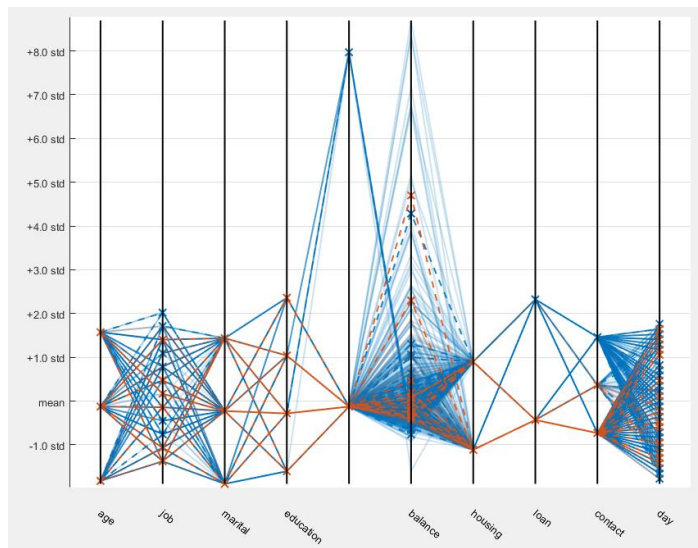The predicted results from the third test obtained are then depicted in the parallel cordinate graph of Figure 4.

Fig. 4.   Third Testing's Parallel Cordinate Graph

V.   CONCLUSION

From the research that has been done can be concluded that the SVM classification algorithm has a good performance in determining the prediction on telemarketing. This study shows that SVM is able to correctly classify data testing. And can be concluded from the results of testing algorithm SVM with 4521 sample data has an average accuracy of 89.16% and error of 10.84%.

From the analysis based on the data can be used by bank to selective choose consumer so there is plenty of time for more potential customers.

REFERENCES

[1]   Kotler P., Keller K. L. (2012). Marketing Management. New Jersey: Prentice Hall

[2]   Tjiptono, Fandy. (2008). Strategi Pemasaran.Yogyakarta: Penerbit Andi

[3]   Saladin, Djaslim. (2006). Manajemen Pemasaran. Bandung: Linda Karya

[4]   Dedić N., Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. Lecture Notes in Business Information Processing. Springer International Publishing. Volume 268, pp. 225-236.

[5]   Hand D., Mannila H., Smyth P. (2001). Principles of Data Mining. Cambrige. MIT Press

[6]   Wu X., et al. (2007). Top 10 Algorithms in Data Mining. New York. Springer

[7]   Vapnik V. (1995). The Nature of Statistical Learning Theory. New York. Springer

[8]   Sulaehani R. (2016). Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination. Jurnal Ilmiah ILKOM Vol. 8 No. 3, Desember 2016

[9]   Kurniawan D., SupriyantoC. (2013). Optimasi Algoritma Support Vector Machine (SVM) Menggunakan Adaboost Untuk Penilaian Risiko Kredit, Jurnal Teknologi Informasi, Vol. 9 No. 1, April 2013

[10] Ispandi, Wahono R. S. (2015). Penerapan Algoritma Genetika untuk Optimasi Parameter pada Support Vector Machine untuk Meningkatkan Prediksi Pemasaran Langsung, Journal of Intelligent Systems, Vol. 1, No. 2, Juni 2015

[11] Sangeeta G., Nisha C. (2012). Data Mining Trend in Past, Current and Future. International Journal of Computing & Business Research.

[12] Larose D. T. (2005). Discovering Knowledge in Data : An Introducing to Data Mining. New Jersey: John Wiley & Sons, Inc.

[13] Han J., Kamber M., Pei J. (2012). Data Mining Concepts and Techniques Third Edition. United Stated of America. Elsevier,

[14] Moro S., Cortez P., Rita P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014