Application of Data Mining with Support Vector Machine (SVM) in Selling Prediction Trend of Spiritual Goods (Case Study: PT. X Bali)

Philipus Novenando Mamang Weking[1*], I Gusti Ngurah Wira Partha[2], and Antonius Ibi Weking[3]

[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3]Department of Electrical and Computer Engineering, Udayana University
*gandamayudarius55@gmail.com

Abstract--- Spiritual activities become an inseparable part of human life. To support these spiritual activities, the need for spiritual equipment that supports the process of running a spiritual activity. PT. X Bali is a company that runs the business of selling spiritual goods. The tight competition and economic factors faced make PT. X Bali wants to predict the sale of goods so that they can see whether the sale of spiritual goods is up or down in order to increase the efficiency. The prediction process can be done with data mining technique using the Support Vector Machine (SVM) method. Data that used for prediction is based on stock data and data on goods sold from the total sales results of the last two years. Based on the results of SVM calculations, the level of prediction accuracy results reached 62.5% and the need for spiritual goods in the following year will be predicted to decline**.**

***Keywords*—Spiritual, Goods, Data Mining, Prediction, Sales, SVM**

I. INTRODUCTION

Spiritual goods are one of the needs used to carry out worship because Indonesia is a country that upholds religion. Catholic is the one of the religions recognized by Indonesia. The church needs the latest spiritual items to renew and replace damaged / lost spiritual items. In fulfilling these needs, the church will usually order spiritual goods through a company in charge of selling spiritual goods. PT. X is one of the companies engaged in the sale of spiritual tools. The high and low sales of spiritual goods are influenced by the quality of the goods, the demand for each religious institution, the economic community, and other supporting factors.

In the process of selling spiritual goods, PT. X Bali faces several problems related to the sale of goods. Economic factors and competition factors in which Indonesia's economic conditions are fluctuating and intense competition causes PT. X Bali must plan a strategy in selling goods. One such strategy is to predict the sale of goods. The prediction is made from the total sales data of goods and stock items that have been available for the past 2 years. To make these predictions, we propose a data mining technique with using Support Vector Machine (SVM) [1] where the purpose of this study is to find out and obtain information on sales predictions of spiritual goods whether up or down for decision making as efficiently as possible to avoid overproduction and reduce the cost of production finance.

II. LITERATURE STUDY

A. Related Works

In this study, several related studies were used as references. Support Vector Machine algorithms can be implemented in several fields. The field of education is one of them, such as predicting the accuracy of student graduation time [1], predictions of drop-out students from schools [2] and drop-out college student prediction from universities [3][4].

SVM can also be implemented in determining the prediction of food commodities such as rice commodities which can determine the prediction of the amount of rice production in the next few years [5]. In the world of economy, many companies have experienced a situation of financial difficulties which triggered bankruptcy. To prevent this, the SVM method was used to predict bankruptcy based on the company's financial statements [6].

B. Data Mining

Data mining is a process that involves searching, grouping, and using data to identify and identify useful information and related knowledge from various large data sources [7]. Data Mining also can be defined as a process of finding patterns and relationships hidden in a large amount of data with the aim of classifying, estimating, forecasting, association rules, sequential pattern, clustering, regression, description and visualization [3]. There are four things are needed in order to effectively data mining such as : high quality data, right of data, examples of which are adequate, and the correct device [11]. In this study, we use forecasting/prediction method.

Prediction/forecasting in data mining is a process of classification of data sets based on variables and values predicted in the future [8].

C. Support Vector Machine

Support Vector Machine is one of the methods used in helping the data mining process in making predictions. SVM is a type of "Supervised Learning" algorithm where SVM has a function to classify data on variables that have predictive value, meaning that these variables have predictable values for the future [4].

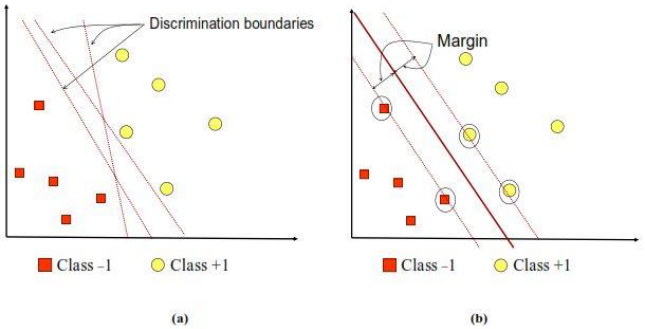

Fig.1 Hyperplane in Support Vector Machine

In SVM, finding the location of the hyperplane is the core of the learning process. The SVM concept can be explained as an effort to find the best hyperplane that functions as a separator of two classes in input space such as Figure 1. In hyperplane, there are only two value such as (+1) which symbolized by red box represent positive value and (-1) which symbolized as yellow circle represent negative value [10].

The best separator hyperplane between the two classes can be found by measuring the margin of the hyperplane and look for the maximum point. Margin is the distance between the hyperplane and the closest pattern of each class. The closest pattern from each class is called support vector. The solid line in figure 1b shows the best hyperplane, which is located right on the middle of the two classes, while the red box and yellow points in the black circle are support vectors [9].

III. Discussion And Results

This research has a process flow of research conducted from the beginning to find the results of the study. The flow of the research methodology as follows in Figure 2:

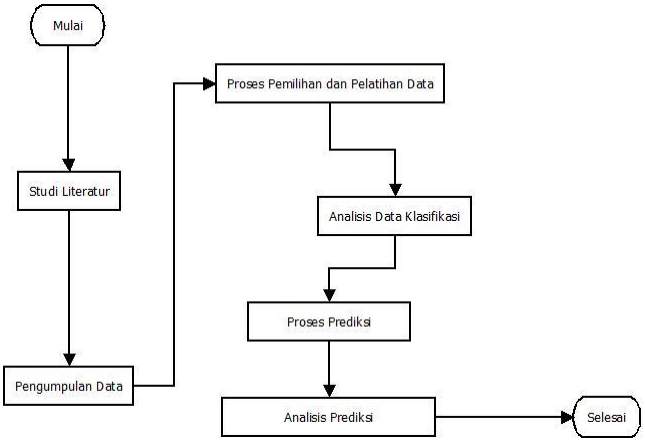

Fig.2 Research Methodology

A. Literature Study

The first step taken in this research is Literature Study. This stage contains a collection of several journals related to research that are used as references.

B. Collecting Data

The next stage is the method of data collection. In this study, the data that we will use in conducting the prediction process is the stock data and goods data sold in the last two years which showed in Table I.

TABLE I
DATASET OF GOODS STOCK AND GOODS SOLD (LAST 2 YEARS)

YEAR	SPIRITUAL GOODS	QTY STOCK	QTY SOLD	PRICE
2017	Lilin	300	161	10000
	Buku Doa Pagi	230	200	15000
	Buku Doa Malam	225	177	10000
	Madah Bakti	220	180	12000
	Dupa	250	200	4000
	Buku Ruah	200	200	5000
	Alkitab	400	150	50000
	Buku Mazmur	270	100	35000
	Rosario	222	150	3500
	Patung Keluarga Kudus	300	250	35000
	Buku Doa Kerahiman	400	289	10000
	Buku Puji Syukur	220	200	35000
	Piala Persembahan	400	140	60000
	Salib	220	170	60000
	Piala Sibori	300	300	55000
	Lampu Pelita	280	222	25000

2018	Lilin	300	161	10000
	Buku Doa Pagi	270	200	15000
	Buku Doa Malam	225	134	10000
	Madah Bakti	230	180	12000
	Dupa	250	200	4000
	Buku Ruah	350	200	5000
	Alkitab	200	172	50000
	Buku Mazmur	270	244	35000
	Rosario	333	300	3500
	Patung Keluarga Kudus	250	250	35000
	Buku Doa Kerahiman	50	100	10000
	Buku Puji Syukur	280	240	35000
	Piala Persembahan	100	140	60000
	Salib	220	170	60000
	Piala Sibori	170	33	55000
	Lampu Pelita	190	222	25000

From data in Table I, to get a better predict results, we simply choose stock data and total of data sold. We simplified those datasets into total data which shown in Table II.

TABLE II
DATASET OF GOODS STOCK AND GOODS SOLD (TOTAL)

SPIRITUAL GOODS	QTY STOCK	QTY SOLD
Lilin	600	322
Buku Doa Pagi	500	400
Buku Doa Malam	450	311
Madah Bakti	450	360
Dupa	500	400
Buku Ruah	550	400
Alkitab	600	322
Buku Mazmur	540	344
Rosario	555	450
Patung Keluarga Kudus	550	500
Buku Doa Kerahiman	450	389
Buku Puji Syukur	500	440
Piala Persembahan	500	280
Salib	440	340
Piala Sibori	470	333
Lampu Pelita	470	444

C. Choose Data and Train Data

The research was continued after dataset is ready. The next research isto finding prediction value with SVM method. In this experiment, the first step to be taken is to determine the data transformation value where the value uses the formula in equation 1

$$X_n = \frac{0{,}8*(X-a)}{b-a} + 0{,}1 \quad (1)$$

Where :
Xn = Value to – n
A = Lowest value
B = Highest value
0.8 and 0.1 = Equation Provision

In the process of data transformation, the item stock variable expressed in (x1) while of the goods sold are expressed in (x2) which range value is about 0-1. The results of the data transformation process are shown in Table III.

TABLE III
DATA TRANSFORMATION RESULTS

X1	X2
0,25	0,9
0,54	0,4
0,21	0,15
0,4	0,15
0,54	0,4
0,54	0,65
0,25	0,9
0,25	0,6
0,33	0,68
0,9	0,65
0,5	0,15
0,7	0,4
0,1	0,4
0,32	0,1
0,8	0,25
0,76	0,25

After getting sample of transformation data, the next step is determining training data. Training data is the data used for training process which to be tested. The training data is taken from X1 and X2 transformation data in Table II from the seventh order, namely (X1 = 0.25, 0.25, 0.33, 0.9) and (X2 = 0.9, 0.6, 0.68, 0.65).The results of the training data are temporary prediction results, by means of the transformation value of X2-X1 as shown in Figure 3.

Data Latih		
0.25	0.9	1
0.25	0.6	1
0.33	0.68	1
0.9	0.65	-1

Fig.3 Results of Training Data

The results from Figure 3 indicate the label of training data based on the results of the reduction in the value of the transformation. If X2-X1value is positive, then the training data label is worth (+1), where as if X2-X1 value is negative, then the training data label is worth (-1). Based on the results of training data that have been obtained and applied to the transformation of other data, it will obtain the results of temporary prediction data as shown in Table IV.

TABLE IV
TEMPORARY PREDICTION RESULTS

X1	X2	Data Results
0,25	0,9	1
0,54	0,4	-1
0,21	0,15	-1
0,4	0,15	-1
0,54	0,4	-1
0,54	0,65	1
0,25	0,9	1
0,25	0,6	1
0,33	0,68	1
0,9	0,65	-1
0,5	0,15	1
0,7	0,4	-1
0,1	0,4	1
0,32	0,1	-1
0,8	0,25	-1
0,76	0,25	-1

D. Classification Data Analysis

In this step, we will determine value of Alpha, Weight and Bias with using Kernel Matrix. From the transformation of existing data in Table III and IV, the next step is to find the kernel matrix value from Equation 2

$$K\ (x, x_i) = \emptyset\ (x_i) \quad (2)$$

The formula in equation 2 is then applied to the results data in Table IV so as to produce the kernel matrix values as in Table V.

TABLE V
KERNEL MATRIX VALUE

2,007	-1,568	-1,431	-1,207
-1,908	2,058	1,207	1,377
4,906	-4,986	1,947	-2,210
3,501	-3,396	-1,634	-1,788

As in Table 4 shows where we get matrix kernel value, then we look for kernel values by summing all the values in the kernel matrix which the results obtained are -0,195.

The next step is to determine the alpha, weight, and bias values in sequence starting from looking for alpha values. Alpha values can be specified in the formula in Equation 3.

$$a1 = a2 = a3 = a4 = a5 = a6 = \frac{6}{\Sigma K(N*N)} \quad (3)$$

The search for alpha values is influenced by the number of kernel matrices tested. In this study, the kernel matrix used is a 4x4 matrix, with the kernel value = -0,195, then the alpha value obtained is = -4,372. Next process is determining weight value and bias value which can be obtained through Equation 4 and 5.

$$w = \sum_{i=1}^{N} a_i y_i K(x, x_i) \quad (4)$$

$$b = -\frac{1}{2}(w.x^{+} + w.x^{-}) \quad (5)$$

Based on the formula in equation 4 which looks for the weight value and equation 5 which looks for the bias value, the weight value is determined by multiplying the alpha value by the value in the kernel matrix while still referring to the training data in Figure 3. The results are in Table VI.

TABLE VI
WEIGHT VALUE RESULTS

-8,77377	6,854645	6,255378	5,276503
8,340894	-8,99672	-5,2765	-6,01967
-21,447	21,79672	-8,51148	9,661202
-15,3049	14,8459	7,143169	7,816393

After getting the weight value which described in Table VI, the bias value will be searched. The bias value can be obtained by selecting the support vector value for class (+1) and class (-1) from the kernel matrix which has a value (+) and value (-) in Table V, then based on formula in Equation 5, the bias value is 17.1441.

E. Prediction Process

Prediction process can be done when the alpha, weight, and bias values have been successfully obtained. Determination of predictive values can be calculated using the formula in Equation 6.

$$f(\phi(x)) = sign(w.\phi(x) + b) = sign(\sum_{i=1}^{N} \alpha_i \gamma_i \phi(x_i)^T . \phi(x) + b) \quad (6)$$

In this study, the value is calculated based on the matrix value, for example taking the matrix value in the first row of the first column (2,007) based on Table IV so that the results are as follows:

=SIGN(((-17.609)+(16.74035)+(-43.0441)+(-30.717)+(17.1441))
= SIGN(-57.4856)
= -1

F. Prediction Analysis

The final step of this research is the application of prediction formulas performed on each value in the kernel matrix in accordance with the SVM method guidelines so that output results are shown in Table VII while the difference of these results are shown in Figure 4.

TABLE VII
PREDICTION RESULTS

X1	X2	Data Results (Temp)	Prediction Results
0,25	0,9	1	-1
0,54	0,4	-1	-1
0,21	0,15	-1	-1
0,4	0,15	-1	1
0,54	0,4	-1	1
0,54	0,65	1	1
0,25	0,9	1	1
0,25	0,6	1	1
0,33	0,68	1	-1
0,9	0,65	-1	-1
0,5	0,15	1	1
0,7	0,4	-1	-1
0,1	0,4	1	-1
0,32	0,1	-1	-1
0,8	0,25	-1	-1
0,76	0,25	-1	-1

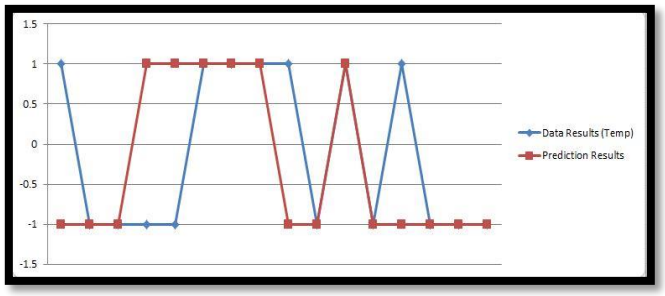

Fig.4 Prediction Results on Graphics

The results can then be determined the level of accuracy based on Figure 4 for using the SVM method with the following calculations:

$$\frac{10}{16} * 100\% = 62{,}50\%$$

IV. CONCLUSION AND SUGGESTIONS

Based on the results of calculations that have been made from the results of classification data using the SVM method, the conclusions that can be drawn from this study are:

1. The level of predictive accuracy obtained using the SVM method is 62.50%.
2. The results of the analysis show that the need for spiritual goods is predicted to decline next year.

In addition to conclusions, there are several suggestions that can be conveyed:

1. The use of variables in SVM must be explained in detail in finding the accuracy of the data calculation process.
2. Requires a combination with other methods that support the data prediction process so that the accuracy of calculations and process results increases.

REFERENCES

[1] Arif Pratama,Randy Cahya Wihandika, and Dian Eka Ratnawati, "Implementasi Algoritma *Support Vector Machine* (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa", Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer : Vol.2, No.4, p.1704-1708, 2018.

[2] Ryci Rahmatil Fiska, "Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera)", Jaringan Sistem Informasi Robotik : Vol.1, No.1, p.42-51, 2017.

[3] Siti Nurhayati, Kusrini, and Emha Taufiq Luthfi, "PREDIKSI MAHASISWA DROP OUT MENGGUNAKAN METODE SUPPORT VECTOR MACHINE", Jurnal Ilmiah SISFOTENIKA : Vol.5, No.1, p.82-93, 2015.

[4] Suprapto, SE., M.Kom., "Penerapan Data Mining Untuk Memprediksi Mahasiswa Drop Out Menggunakan Support Vector Machine", KOMPUTAKI : Vol.1, No.1,p.14-29, 2015.

[5] Atik Nurmasani, Ema Utami, and Hanif Al Fatta, "Analisis *Support Vector Machine* Pada Prediksi Produksi Komoditi Padi" Jurnal Informasi Interaktif : Vol. 2, No. 1, p.39-46, 2017.

[6] Lestari Handayani, and Fitriandini, "PREDIKSI KEBANGKRUTAN PERUSAHAAN MENGGUNAKAN SUPPORT VECTOR MACHINE", Jurnal IJECE Vol. VII No. 1, 2016.

[7] Turban, E., "Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1", Penerbit Andi : Yogyakarta, 2005.

[8] Larose, Daniel T., "Discovering Knowledge in Data : An Introduction to Data Mining",

[9] Ivo Colanus Rally Drajana, "METODE *SUPPORT VECTOR MACHINE* DAN *FORWARD SELECTION* PREDIKSI PEMBAYARAN PEMBELIAN BAHAN BAKU KOPPA", ILKOM Jurnal Ilmiah : Vol. 9 No. 2, p. 116-123, 2016.

[10] Anto Satriyo Nugroho, Arief Budi Witarto, and Dwi Handoko "Support Vector Machine-Teori dan Aplikasinya dalam Bio informatika", Kuliah Umum IlmuKomputer.com, 2003.

[11] Hendrayana Putu Bagus Surya, Rifky Lana Rahardian, and Made Sudarma, "Application of Neural Network Overview In Data Mining", International Journal of Engineering and Emerging Technology, Vol. 2, No. 1, p. 94-96, 2017.