Classification of Credit Card Default Clients Using LS-SVM Ensemble

Armin Lawi
Department of Computer Science
Hasanuddin University
Makassar, Indonesia
armin@unhas.ac.id

Firman Aziz
Faculty of Mathematics and Natural Sciences
Pancasakti University
Makassar, Indonesia
firmanaziz88@gmail.com

***Abstract*— Finding knowledge from a database and turning it into useful information is a big challenge. The use of machine learning helps analyze data and contribute to delivering results that can be acted upon by the company. SVM is one of machine learning method that has better performance than other machine learning method but sensitive to parameter setting and training sample. the performance accuracy of the SVM method can be improved using the LS-SVM and Ensemble method. This research proposes the LS-SVM ensemble to identify the prospective credit cards client that will default. The Least Square SVM ensemble method has the highest percentage with a difference of 1.7% from SVM and 0.6% from Least Square SVM.**

Keywords—Credit Card, SVM, LS-SVM, AdaBoost, Ensemble

I. INTRODUCTION

A credit card is a physical card that is used easily to pay the amount of a shopping bill. Cardholders may use it to provide a promise of payment for the cost of services and goods [1]. Many banks offer credit card facilities to potential customers with diverse promotions that provide benefits. But in recent years there have been problems when credit card holders were unable to make payments and accumulated credit debt increased. The crisis caused a decrease in consumer confidence (whether the consumer was able to pay the bills or not). Finding useful knowledge from the database and turning information into actionable results is a big challenge facing the company.

To overcome these problems, credit card risk prediction should be made using the data history of credit card customers. Credit card customer history data will be classified to determine the possibility to pay for the credit or not. To classify the data set used methods of data mining.

Data mining is a computational process that shows patterns in a data set using methods such as artificial intelligence, machine learning, statistics etc [2]. While [3] Data mining is a process of exploration and analysis, by means of automatic or semi-automated, with large amounts of data to find meaningful patterns and rules.

Currently, data mining is an indispensable tool in decision support systems and plays a role in market segmentation, customer service, fraud detection, credit ratings, behavior and benchmarking [4]; [5].

Research [6] suggests that estimating probability would be better than classifying customers into binary results as it represents the real probability of a real default. But with the evolution of artificial intelligence and machine learning, the classification is also used to predict credit card risk [7]; [5].

Several classification methods have been used in the case of credit risk, research [8] applies the BayesNet, Stacking, NaiveBayes, Random Forest, and Random Tree methods to see the performance of each in classifying credit risk.[9] apply data mining methods such as FLDA, J48, Logistic Regression, Naive Bayes, MLP, and IBK. the results show J48 and Logistic Regression show good predictive accuracy. whereas [10] apply estimates to the Multilayer Perceptron (MLP) and k-Nearest Neighbors (KNN) and successfully apply in the case of credit risk.

This study proposes other data mining methods that are often used in the classification, namely SVM to identify potential credit card customers who will fail. SVM is a machine learning method that has better performance than other machine learning methods but is sensitive to parameter settings and training samples. to overcome this, this study also improved the performance of the SVM method using the LS-SVM method [11]; [12]; [13] and the Ensemble method [14]; [15]; [16].

The Ensemble Method concept combines several sets of models that solve similar problems to get a more accurate model [17]. one of the ensemble methods is Adaptive Boosting.

II. LITERATURE REVIEW

A. Classification using Support Vector Machine

SVM is one of the methods in the problem of pattern classification. The theory underlying SVM has developed since the 1960s, but was only introduced by Vapnik in 1995 [18]. SVM finds the optimal hyperplane that categorizes training data input into two classes (good and bad). Figure 1. shows the basic support vector machine [18].

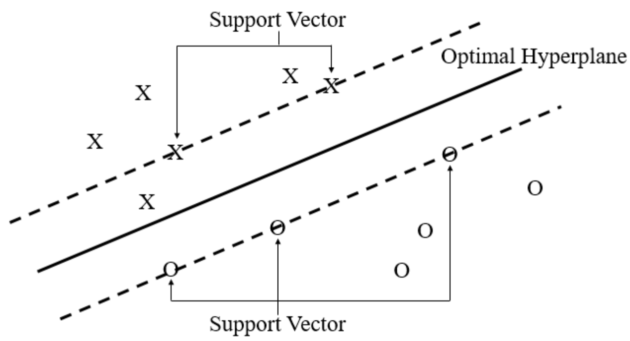

Fig. 1. Base of Support Vector Machines

B. *Classification using Least Square Support Vector Machine*

LS-SVM is a modified form of SVM which was first introduced by Suykens and Vandewalle in 1999 [19]. LS-SVM forms a bounding field that satisfies the following equation:

$$x_i.\boldsymbol{w}^T + b = +1 \; for \; y_i = +1,$$

$$x_i.w^T + b = -1 \text{ for } y_i = -1 \tag{1}$$

Similar to SVM, LS-SVM can also be used to classify data that can not be separated linearly by the use of kernel functions.

In contrast to the use of soft margin SVM, LS-SVM using quadratic hyperplane. Then the form of the LS-SVM is as follows:

$$\min \frac{1}{2}\|w\|^2 + \frac{c}{2}\|\xi_i\|^2,$$

With constraints

$$y_i(\varphi(x_i).\boldsymbol{w}^T + b) = 1 - \xi_i, k = 1,2,\dots,N \tag{2}$$

ξ is the slack variable that determines the classification error rate, C>0 is the parameter that determines the penalty due to errors in the data classification and K is the kernel function used.

C. *Classification Ensemble*

The ensemble method can reduce classification errors effectively, and is believed to perform well compared to the use of a single classifier. The main idea of the ensemble method is to combine several sets of models that solve a similar problem to obtain a more accurate model [11]. Compared to individual classifiers, they only learn and train just one data set, but ensemble classifiers learn and train a variety of data generated from the original data set. then the result will build a set of hypotheses from the trained data and produce better accuracy [20].

Some ensemble classification techniques have been developed such as bagging, boosting, random forest and rotation forest but boosting to be used in this study.

Boosting is a common and effective method for building accurate classifiers by combining weak classifiers. The use of boosting is preferred because it focuses on misclassified issues and has a tendency of increased accuracy compared to the bagging method.

The focus of this method is to generate a series of base classifiers. The training sets used for each base classifiers are selected based on the performance of the previous classifiers. In the boosting, the sample is not predicted correctly by the classifiers in the series will be selected more often than the samples that had been predicted correctly. Thus, boosting tries to produce a new base classifier that are better for predicting samples that in the previous base classifiers have poor performance [20]. One of the most popular algorithms of the boosting method is the AdaBoost algorithm.

AdaBoost is an ensemble method (meta-learning) that builds a classifier in repetitive mode. In each iteration, it will call a simple learning algorithm (called the learning base) which returns to a classifier and sets the coefficient weight [21].

The weighting technique in the AdaBoost algorithm is done using the following algorithm [22]:

1. Input: Training data along with its label $\{(x_1, y_1), \dots, (x_N, y_N)\}$, Component Learner, And the number of iterations of T.
2. Initialization of training data weight

$$D_i^1 = \frac{1}{N}, i = 1, \dots, N \tag{3}$$

3. For iteration t=1,… ,T
   a. Counting classification errors on h_t

$$\varepsilon_t = \sum_{i=1}^{N} D_i^t (y_i \neq h_t(x_i)) \tag{4}$$

   $h_t(x_i)$ is a model obtained in the distribution D_t.

   b. The trust learning index is calculated as:

$$\alpha_t = \frac{1}{2}\ln\left(\frac{(1-\varepsilon_t)}{\varepsilon_t}\right) \tag{5}$$

   c. Update the training sample weight

$$D_i^{t+1} = \frac{D_i^t \exp(\alpha_t \times h_t(x_i) \neq y_i)}{\sum_{i=1}^{N} D_i^t} \tag{6}$$

   d. Normalization D_i^{t+1} so the number becomes one.

4. Last learning output

   Combination of all classifications

$$H = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x_i)\right) \tag{7}$$

D. *Performance Evaluation*

The calculation of the classification accuracy value is obtained using the following confusion matrices [23].

TABLE I. CONFUSION MATRIX/ CONTINGENCY TABLE

Actual/Prediction	Good Clients	Bad Clients
Good Clients	TP	FN
Bad Clients	FP	TN

The performance results of each classification are evaluated based on 3 measurements, namely: Accuracy, Sensitivity, and Specificity. The measurement uses the equation as follows:

$$accuracy = \frac{TP+TN}{TN+FP+FN+TP} \tag{8}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

$$Specificity = \frac{TN}{TN+FP} \tag{10}$$

III. EXPERIMENTAL

A. *Dataset*

In this study using benchmarking data of Taiwan bank credit card clients. Using default binary binary variable (Yes = 1, No = 0), as response variable. Consisting of 23 attributes and obtained from UCI Machine Learning Repository. The description of the dataset attribute is as follows:

- X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.

- X2: Gender (1 = male; 2 = female).
- X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
- X4: Marital status (1 = married; 2 = single; 3 = others).
- X5: Age (year).
- X6 - X11: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows: X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .; X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
- X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2005; X13 = amount of bill statement in August, 2005; . . .; X17 = amount of bill statement in April, 2005.
- X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .; X23 = amount paid in April, 2005.

B. Normalization Data

The attributes of 'education' and 'marital_status' have undefined values, so the two attributes will be done by cleaning the data to re-encode the undefined values into other categories. furthermore, data reduction is done by reducing the size of the data set to achieve the same class representation between the default and non-default classes. The dataset is subtracted from 30,000 to 13,270 with the representation of each class of 6,635.

C. Implementation

The focus of this study was to look at the performance of the LS-SVM ensemble algorithm as well as to increase the accuracy of the SVM and LS-SVM single classification. The data will be partitioned into two sets of 70% training set and 30% testing set. The results of this separation contain 9,290 training data and 3,980 test data. Then apply the SVM algorithm, LS-SVM, and LS-SVM ensemble.

IV. Result

After the testing process using credit card client data from Taiwan bank obtained results as in Table II.

Table II shows the overall results obtained, Least Square SVM ensemble can improve the accuracy of classification prediction or reduce classification errors in SVM and Least Square SVM. The accuracy of classification using SVM is 66.9% and Least Square SVM is 68%. When using Least Square SVM ensemble the classification accuracy changed to 68.6%.

The Least Square SVM ensemble method has the highest percentage with a difference of 1.7% from SVM and 0.6% from Least Square SVM. This proves that the Least Square SVM ensemble method is better than other methods.

TABLE II. Confusion Matrix/ Contingency Table

Classification	**Performance measurement**		
	accuracy	***Sensitivity***	***Specificity***
SVM	66.9%	65.5	68.5
LS-SVM	68%	67.9%	68.1
LS-SVM Ensemble	68.6%	67.9%	69.5

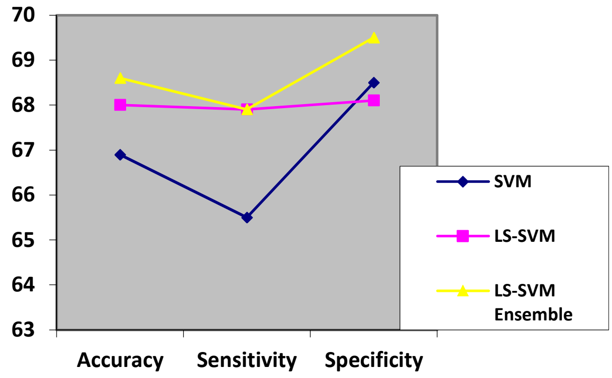

Fig. 2. Performance Algorithm

A. Comparison of SVM with LS-SVM

The sensitivity value is 67.9%, which means that the ability to provide positive results for prospective credit card customers default is 67.9%. Specificity value is 68.1%, which means the ability to give negative results to prospective credit card customers who will not default to 68.1%. Accuracy score of 68%, which means the ability to correctly detect all tested subjects of 68%.

Measurements with a high level of sensitivity are needed to detect potential customers who will default. A higher specificity is needed to reinforce the alleged default potential customers. Although the value of high sensitivity and specificity is lower than the SVM method. Least Square SVM has better performance with an accuracy difference of 1.7%.

B. Comparison of LS-SVM with LS-SVM ensemble

From the results of experiments performed, by ensemble using algorithms AdaBoost performance of the method Least Square SVM ensemble can be improved. This is because of the algorithm AdaBoost re-weighting against the wrongly classified data and retraining each iteration to the specified iteration limit. However, the more iterations performed by the Adaboost algorithm, it will not guarantee the performance of the component learner will be better.

The Least Square SVM ensemble method has a high accuracy and specificity value compared to Least Square SVM and the same sensitivity value. This shows that the Least Square SVM ensemble is effective enough to detect potential credit card customers who will default. The Least Square SVM ensemble method is also more flexible for the various data sharing models between training data and test data.

V. Conclusion

This study proposes the SVM method and improves the performance of the SVM method using the LS-SVM method and the Ensemble method to identify potential credit card customers who will fail. the results show that the Least Square

SVM ensemble can improve the accuracy of classification predictions or reduce misclassification in SVM and Least Square SVM. The Least Square SVM ensemble method has the highest percentage with a difference of 1.7% from SVM and 0.6% from Least Square SVM. This proves that the Least Square SVM ensemble method is better than other methods.

Although the sensitivity and high specificity values are lower than the SVM method. Least Square SVM has a better performance with a difference in accuracy of 1.7%. The Least Square SVM ensemble method is effective enough to detect potential credit card customers who will fail to pay with high accuracy, sensitivity and specificity compared to SVM and LS-SVM methods. The Least Square SVM ensemble method is also more flexible for various data sharing models between training data and test data.

REFERENCES

[1] O'sullivan, Arthur, and Steven M. Sheffrin. "Economics: Principles in action", 2003.

[2] Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." *IEEE Transactions on Knowledge and data Engineering 8.6*, 1996: 866-883.

[3] Berry, Michael, and Gordon Linoff. *Mastering data mining: The art and science of customer relationship management*. John Wiley & Sons, Inc., 1999.

[4] Giudici, Paolo. "Bayesian data mining, with application to benchmarking and credit scoring". *Applied Stochastic Models in Business and Industry* 17.1, 2001: 69-81.

[5] Thomas, Lyn C. "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers". *International journal of forecasting* 16.2, 2000: 149-172.

[6] Yeh, I-Cheng, and Che-hui Lien. "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients". *Expert Systems with Applications* 36.2, 2009: 2473-2480.

[7] Koh, Hian Chye, and Chan Kin Leong Gerry. "Data mining and customer relationship marketing in the banking industry". *Singapore Management Review* 24.2, 2002: 1.

[8] Venkatesh, Ajay, and Shomona Gracia Jacob. "Prediction of credit-card defaulters: a comparative study on performance of classifiers." *International Journal of Computer Applications*145.7 (2016).

[9] Pasha, M., Fatima, M., Dogar, A. M., & Shahzad, F. "Performance comparison of data mining algorithms for the predictive accuracy of credit card defaulters." *Int. J. Comput. Sci. Netw. Secur* 17.3 (2017): 178-183.

[10] KOKLU, Murat, and Kadir SABANCI. "Estimation of Credit Card Customers Payment Status by Using kNN and MLP." *Int. J. Intell. Syst. Appl. Eng* 4 (2016): 249-251.

[11] Zhou, Ligang, Kin Keung Lai, and Lean Yu. "Least squares support vector machines ensemble models for credit scoring". *Expert Systems with Applications* 37.1, 2010: 127-133.

[12] Jafar, Nurkamila, Sri Astuti Thamrin, and Armin Lawi. "Multiclass classification using Least Squares Support Vector Machine." Computational Intelligence and Cybernetics (CYBERNETICSCOM), 2016 International Conference on. IEEE, 2016.

[13] Lawi, Armin, and Yudhi Adhitya. "Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine." Journal of Physics: Conference Series. Vol. 979. No. 1. IOP Publishing, 2018.

[14] Feng, Guang, Jia-Dong Zhang, and Stephen Shaoyi Liao. "A novel method for combining Bayesian networks, theoretical analysis, and its applications". *Pattern Recognition* 47.5, 2014: 2057-2069.

[15] Lawi, Armin, Firman Aziz, and Syafruddin Syarif. "Ensemble GradientBoost for increasing classification accuracy of credit scoring." Computer Applications and Information Processing Technology (CAIPT), 2017 4th International Conference on. IEEE, 2017.

[16] Lawi, Armin, Ali Akbar Velayaty, and Zahir Zainuddin. "On identifying potential direct marketing consumers using adaptive boosted support vector machine". Computer Applications and Information Processing Technology (CAIPT), 2017 4th International Conference on. IEEE, 2017.

[17] Zhang, Cha, and Yunqian Ma, eds. *Ensemble machine learning: methods and applications*. Springer Science & Business Media, 2012.

[18] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks". *Machine learning* 20.3, 1995: 273-297.

[19] Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers". *Neural processing letters* 9.3, 1999: 293-300.

[20] Kuncheva, Ludmila I. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.

[21] Kégl, Balázs, and Róbert Busa-Fekete. "Boosting products of base classifiers". *Proceedings of the 26th Annual International Conference on Machine Learning. ACM,* 2009.

[22] Freund, Yoav, Robert Schapire, and Naoki Abe. "A short introduction to boosting". *Journal-Japanese Society For Artificial Intelligence* 14.771-780, 1999: 1612.

[23] Provost, Foster, and Ron Kohavi. "Guest editors' introduction: On applied research in machine learning". *Machine learning* 30.2, 1998: 127-132.