Long-term deposits prediction: a comparative framework of classification model for predict the success of bank telemarketing

Ahmad Ilham*1, Laelatul Khikmah2, Indra3, Ulumuddin4 and Ida Bagus Ary Indra Iswara5

1Informatics Department, Universitas Muhammadiyah Semarang, Semarang, Indonesia
2Statistics Department, Akademi Statistika Muhammadiyah, Semarang, Indonesia
3Engineering Department, Universitas Sulawesi Barat, Majene, Indonesia
4Informatics Engineering Department, STIMIK YMI Tegal, Tegal, Indonesia
5Informatics Department, STMIK STIKOM Indonesia, Denpasar, Bali

*ahmadilham@unimus.ac.id

Abstract. The long-term deposits product is often offered to prospective customers by the bank through telemarketing strategy. How to maximize customer value through telemarketing strategy is a major focus in this research. Therefore, required a model that can classify potential customers with the potential to increase corporate earnings. The Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), K-Nearest Neighbour (K-NN), Support Vector Machine (SVM), Neural Network (NN), and Logistic Regression (LR) model have been proposed to compare. The comparison of these algorithms evaluated using the Protestant Bank dataset of the UCI Machine Learning repository, for performance algorithms using Area Under Curve (AUC) and Accuracy. From the experiment, the results show that SVM yield promising results with the Accuracy and AUC 97.07% and 0.925 respectively. It can be concluded that SVM is the best choice for classifying prospective customers who have the potential to be interested in time deposit products that are offered by telephone or cellular as distinguished from other classification algorithms.

1. Introduction

The product promotions in marketing activities is a strategy that is often used to increase corporate earnings. One of the promotional strategies that are usually carried out by retail banks to offer bank products is telemarketing strategies. The strategy is a direct marketing strategy to the customer by contacting them to meet certain goals. By focusing on interaction with customers using remote communication on a contact center (CC) will facilitate the promotion operations management section [1].

Building a telemarketing business strategy is necessary to maximize customer value through the evaluation of the information held, so that as to build a longer and closer relationship in appropriate with business needs. According Neysiani et al [2] to achieve this, it is required to classify the criteria or characteristics of prospective customers, if this is not done, it can lead to inefficiencies in the
strategy. For example, a bank offers deposit products to a prospective customer without classifying a potential customer, as a result, many of the prospective customers when interacted by the agent over the phone provide various responses such as those interested in the offer, others are not interested.

Based on this information, it is required a model that can classify potential customers who have potential to be interested in products offered by the bank. There are several classification algorithm to be used in this case, namely Decision Tree (DT) [3], Naïve Bayes (NB) [4], K-Nearest Neighbor (K-NN), Random Forest (RF) [5], Support Vector Machine (SVM) [6], Neural Network (NN), and Logistic Regression (LR) [5], [7].

One of the many researchers has implemented and compiled four algorithms of data mining classification such as LR, DT, NN and SVM in the field of Decision Support System (DSS) to predict the success of telemarketing business strategies in banks [1]. The results of their study indicate that NN is the best algorithm with Area Under Curve (AUC) amount of 0.832. However, their study did not perform tests that show the best algorithm based on Accuracy. Therefore, in this paper will be discussed about testing Accuracy by using Accuracy and AUC. For statistical difference test will use another statistical test to find out which algorithm has the most accurate performance.

In this study This paper is organized as follows. In section 2, the methodology of this study is explained including the proposed algorithm. The experimental results and discussion of comparing the proposed algorithm with other prior researches are presented in section 4. Finally, our work of this paper is summarized in the last section.

2. Proposed comparison algorithm

The proposed comparison algorithm is shown in Figure 1. The comparison algorithm consists of a dataset, a classification algorithm, a model validation, a model evaluation and a model comparison. The purpose of this study is to find the best algorithm that is come strong performance of long-term deposits prediction. Therefore it can help policymakers.

2.1. Dataset

The proposed algorithm evaluated using Bank Portugal dataset from UCI machine learning where this dataset belongs from bank retail in Portugal by 2008-2010 years. Table 1, shown the dataset description consists of 4120 records, 20 attributes and 2 labels is deposit and non-deposit. Figure 1, shows the proposed comparison of classification algorithms for long-term deposit prediction.

2.2. Classification algorithm and model validation

As shown in Figure 1, we use cross-validation to estimate the statistical performance of a learning algorithm. In this study, we use 10 fold which will split into 10 parts dataset, 1 part as a testing dataset and the rest as training datasets and this process repeated 10 times. 7 classifiers have been selected such as Nearest Neighbors (KNN), Decision Tree (DT, RF), Neural Networks (NN), the statistical classifier (LR, NB) and Support Vector Machine (SVM). This comparative selection aims to achieve a balance between the classification algorithms used in long-term deposit prediction. After learning algorithms process complete, all of the algorithms will feed with the testing dataset and then we record the evaluation result.

2.3. Model evaluation

To evaluate the performance of classifiers model, we apply Area Under Curve (AUC) as an indicator Accuracy in our experiments. Lesman et al [8] suggest the use of AUC to increase the comparability of cross-study in which AUC has the potential to significantly increase convergence across empirical experiments in long-term deposit prediction. In addition, AUC has a clear statistical interpretation. As stated by Gorunescu [9], basic guidelines for classifying the Accuracy of diagnostic tests based on AUC can be seen in Table 2.
Table 1. Dataset Descriptions

| Cat. | # | Atribution | Type  |
|------|---|------------|-------|
| Bank client data of the last contact | 1 | Age | * |
| 2 | Job type | ** |
| 3 | Marital status | ** |
| 4 | Education | ** |
| 5 | Default: Has the credit in default? | ** |
| 6 | Housing: Has a housing loan? | ** |
| 7 | Loan: Has personal loan? | ** |
| Other attributes | 8 | Contact: Contact communication type | ** |
| 9 | Month: Last contact month of year | ** |
| 10 | Day of wee: Last contact day of the week | ** |
| 11 | Duration: Last contact duration | * |
| Social and economic context attributes | 12 | Campaign: number of contacts performed during this campaign and for this client | * |
| 13 | Pdays: Number of days that passed by after the client was last contacted from a previous campaign | * |
| 14 | Previous: Number of contacts performed before this campaign and for this client | * |
| 15 | Poutcome: Outcome of the previous marketing campaign | ** |
| 16 | Emp.var.rate: Employment variation rate-quarterly indicator | * |
| 17 | Cons.price.idx: Consumer price index – monthly indicator | * |
| 18 | Cons.conf.idx: Consumer confidence index-monthly indicator | * |
| 19 | Euribor3m: Euribor 3 month rate-daily indicator | * |
| 20 | Nr.employed: Number of employees-quarterly indicator | * |
| 21 | Result label: Has the client subscribed a term deposit? | *** |

Numeric, **Categorical, * Binary

Table 2. Confusion matrix.

| AUC          | Meaning               |
|--------------|-----------------------|
| 0.90-100     | Excellent classification |
| 0.80-0.90    | Good classification    |
| 0.70-0.80    | Fair classification    |
| 0.60-0.70    | Poor classification    |
| < 0.60       | Failure                |

2.4. Model comparison
After getting the Accuracy of each model tested, then test the difference by using T-test. According to Dietterich [10] that using T-test based on random subsampling has high probability error of A
algorithm which wrongly detects the difference when there is no difference. The results in this test are based on 7 iterations of 10-fold cross-validation. The $7 \times 10$cv on T-test following Eq (1),

$$
\beta_{c} = \frac{AC^{(1)}_{\text{c}}}{\sqrt{\frac{1}{\text{7}} \sum_{i=1}^{\text{10}} r^{(i)}}},
$$

where $AC^{(1)}_{\text{c}}$ is the difference in Accuracy from the first fold of the first replication of 10-fold cross-validation and $r^{(i)}$ is the variance computed from the $i$-th replication.

3. Experiment results and discussion

The experiments are tested using computing platform based on Intel Celeron 2.16 GHz CPU, 8 GB RAM and Microsoft Windows 7 64-bit operating system, and Rapid Miner Studio 8.2 as data analytics tool. RapidMiner will be used to measure Accuracy (ACC), AUC and confusion matrix.

First of all, we conducted an experiment on Bank Portugal dataset by using 7 classifications algorithms. Table 3, shown Accuracy and AUC evaluation results of 7 classification algorithms with the same datasets. Figure 2 and 3 shown a bar chart by Accuracy and AUC results of all algorithm of long-term deposit prediction.

**Table 3.** Accuracy and AUC of 7 classification algorithms.

| Performance | DT | NB | RF | K-NN | SVM | NN | LR |
|-------------|----|----|----|------|-----|----|----|
| Accuracy    | 90.00 | 87.18 | 89.05 | 88.23 | 91.07 | 88.59 | 89.05 |
| AUC         | 0.645 | 0.868 | 0.744 | 0.500 | 0.925 | 0.858 | 0.218 |

**Figure 2.** Accuracy results of all algorithms on long-term deposits prediction.

**Figure 3.** AUC results of all algorithms on long-term deposits prediction.

As a seen in Table 3, on Accuracy evaluate the SVM algorithm was the first superior algorithm with high Accuracy 91.07%, followed by DT 90.00%, RF 89.05%, LR 89.05%, NN 88.59%, K-NN 88.23% and the last NB 87.18%. While for Area Under Curve (AUC) evaluation, SVM is also superior to other comparative algorithms with the AUC values of SVM is 0.925, followed by NB 0.868, then NN 0.858, RF 0.744, DT 0.645, K-NN 0.500 and LR 0.218. We can see the different results between Accuracy and AUC, NB shows a very significant change from the Accuracy evaluation to AUC. In this study will further review the results of AUC. This conclusion is based on advice by Lesman et al [8] and Wahono et al. [11] in software prediction that the AUC has the
potential to significantly increase convergence across empirical experiments as it separates predictive performance from operating conditions and represents a predictive measure of general. In this comparison, we also used a t-test as a statistical significance test as shown in Table 4.

### Table 4. T-test AUC comparison with 7 classification algorithms

|       | DT       | NB       | RF       | K-NN     | SVM      | NN       | LR       |
|-------|----------|----------|----------|----------|----------|----------|----------|
| DT    | 0.900 +/- 0.012 | 0.872 +/- 0.015 | 0.891 +/- 0.002 | 0.882 +/- 0.009 | 0.911 +/- 0.013 | 0.886 +/- 0.009 | 0.891 +/- 0.001 |
| NB    | 0.872 +/- 0.015 | 0.000 | 0.020 | 0.001 | 0.067 | 0.035 | 0.019 |
| RF    | 0.891 +/- 0.002 | 0.001 | 0.071 | 0.000 | 0.146 | 0.072 | 0.001 |
| K-NN  | 0.882 +/- 0.009 | 0.013 | 0.000 | 0.486 | 1.000 |
| SVM   | 0.911 +/- 0.013 | 0.000 | 0.486 | 0.012 |
| NN    | 0.886 +/- 0.009 | 0.002 | 0.000 | 0.136 |
| LR    | 0.891 +/- 0.001 |

We used t-test to detect whether there are differences between with other algorithms and also to show which algorithm is promising performance. The significant difference between the actual average value in AUC result is highlighted with boldface print, meaning its value is less than alpha = 0.050. It is known that the order of the best classification algorithms to the lowest is the first SVM, the second DT, the third RF, the fourth LR, the fifth NN, the sixth K-NN and the last is NB. However, between SVM and DT there is no significant difference, between RF, LR and NN there is no significant difference, between NN, K-NN and NB there is no significant difference, and between K-NN and NB there is no difference which is significant (Table 4). Specifically for NB algorithm, it cannot be concluded that the NB algorithm is not superior because the performance of NB algorithm can still be improved by weighting technique feature as feature selection algorithm [12], [13]. Based on t-test, SVM algorithm indicated excellent and competitive result with the state-of-the-art traditional algorithms result.

4. Summary

Required a model to find out the information needed to find out products provided by the bank, either by phone or cellular. Therefore, a model is required to construct these potential customers. The classification algorithms used are DT, NB, RF, KNN, SVM, NN and LR.

The experimental results show that the best algorithm is SVM, second DT, third RF, fourth LR, fifth NN, seventh k-NN and the last is NB. However, there is no significant difference between SVM and DT, between RF, LR and NN there is no significant difference, between NN, K-NN and NB there is no significant difference, and between K-NN and NB there is no significant difference.

In this study, it is known that SVM is the best choice for classifying prospective customers who have the potential to be interested in time deposit products that are offered by telephone or cellular as distinguished from other classification algorithms. However, this paper does not do preprocessing or pay attention to the features or features in the dataset, only directly use a dataset that is ready from the UCI repository. Therefore, it is possible that the test results will be different if determining based on preprocessing the dataset before do the classification model is applied to be tested.
References

[1] Moro S and Cortez P and and Rita P, 2014 A data-driven approach to predict the success of bank telemarketing *Decis. Support Syst.* 62, February 2014 p. 22–31.

[2] Neysiani B S Soltani N and Ghezelbash S, 2015 A framework for improving find best marketing targets using a hybrid genetic algorithm and neural networks in *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*.

[3] Hoedvik V et al., 2015 A Framework for Improving Find Best Marketing Targets Using a Hybrid Genetic Algorithm and Neural Networks *2015 2nd Int. Conf. Knowledge-Based Eng. Innov.* p. 733–738.

[4] Vedala R and Kumar B R, 2012 An application of Naive Bayes classification for credit scoring in e-lending platform *Proc. - 2012 Int. Conf. Data Sci. Eng.* ICDSE 2012 p. 81–84.

[5] Asare-Frempong J and Jayabal M, 2017 Predicting customer response to bank direct telemarketing campaign *2017 Int. Conf. Eng. Technol. Technopreneurship, ICE2T 2017* January p. 1–4.

[6] Zhen W and Wenjuan S, 2016 Commercial Bank Credit Risk Assessment Method Based on Improved SVM *2016 Int. Conf. Intell. Transp. Big Data Smart City* p. 353–356.

[7] Khikmah L Wijayanto H and Syafitri U D, 1998 Modeling Governance KB with CATPCA to Overcome Multicollinearity in the Logistic Regression *Braz. Dent. J.* 9, 1 p. 3–10.

[8] Lessmann S Baesens B Mues C and Pietsch S, Jul. 2008 Benchmarking Classification Models for Software Defect Prediction: A Proposed Framework and Novel Findings *IEEE Trans. Softw. Eng.* 34, 4 p. 485–496.

[9] Gürunescu F, 2011 *Data Mining: Concepts, Models and Techniques*.

[10] Dietterich T G, 1998 Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms *Neural Comput.*

[11] Wahono R S Herman N S and Ahmad S, 2014 A comparison framework of classification models for software defect prediction *Adv. Sci. Lett.* 20, 10–12 p. 1945–1950.

[12] Asmono R T Wahono R S and Syukur A, 2015 Absolute Correlation Weighted Naïve Bayes for Software Defect Prediction *J. Softw. Eng.* 1, 1 p. 38–45.

[13] Taheri S Yearwood J Mammadov M and Seifollahi S, 2014 Attribute weighted Naïve Bayes classifier using a local optimization *Neural Comput. Appl.* 24, 5 p. 995–1002.