

TELEMARKETING BANK SUCCESS PREDICTION USING MULTILAYER PERCEPTRON (MLP) ALGORITHM WITH RESAMPLING

Siti Masturoh^{1*}; Fitra Septia Nugraha²; Siti Nurlela³; M. Rangga Ramadhan Saelan⁴;
Daniati Uki Eka Saputri⁵; Ridan Nurfalah⁶;

^{1, 3, 4, 5, 6} Program Studi Sistem Informasi

²Program Studi Teknik Informatika

STMIK Nusa Mandiri

www.nusamandiri.ac.id

siti.uro@nusamandiri.ac.id (*); fitra.fig@nusamandiri.ac.id; siti.sie@nusamandiri.ac.id;

rangga.mgg@nusamandiri.ac.id; daniati.due@nusamandiri.ac.id; ridan.rlh@nusamandiri.ac.id

(*) Corresponding Author

Abstract—Telemarketing is a promotion that is considered effective for promoting a product to consumers by telephone, other than that telemarketing is easier to accept because of its direct nature of offering products to consumers. Telemarketing is also considered to help increase a company's revenue. The problem of predicting the success of a bank's telemarketing data must be done using machine learning techniques. Machine learning used in the available historical data is a bank dataset of 45211 instances at 17 features using the multilayer perceptron algorithm (MLP) with resampling. The use of resampling aims to balance the unbalanced data resulting in an accuracy value of 90.18% and a ROC of 0.89%. Meanwhile, if the data resampling is not used in the multilayer perceptron (MLP) algorithm, the accuracy value is 88.6 and ROC is 0.88%. The use of resampling data becomes more effective and results in higher accuracy values.

Keywords: Telemarketing; Multilayer Perceptron (MLP); Resampling.

Abstrak—Telemarketing merupakan salah satu promosi yang dianggap efektif untuk melakukan sebuah promosi produk kepada konsumen melalui telepon, selain itu telemarketing lebih mudah untuk diterima karena sifatnya langsung menawarkan produk kepada konsumen. Telemarketing juga dianggap membantu untuk meningkatkan pendapatan sebuah perusahaan. Permasalahan mengenai prediksi keberhasilan sebuah data telemarketing bank harus di lakukan dengan teknik machine learning. Machine learning yang digunakan pada data history yang tersedia yaitu dataset bank sebanyak 45211 instance pada 17 feacture menggunakan algoritma multilayer perceptron (MLP) dengan resampling. Penggunaan resampling bertujuan untuk menyeimbangkan data yang tidak seimbang sehingga menghasilkan nilai akurasi

sebesar 94,27 % dan ROC sebesar 0,89 %. Sedangkan jika tidak menggunakan resampling data pada algoritma multilayer perceptron (MLP) menghasilkan nilai akurasi 90,18 dan ROC sebesar 0,89 %. Penggunaan resampling data menjadi lebih efektif dan menghasilkan nilai akurasi lebih tinggi.

Kata Kunci: Telemarketing, multilayer perceptron (MLP), resampling.

INTRODUCTION

Telemarketing is one of the promotions that can be considered effective for marketing or introducing products by telephone. Also, telemarketing tends to be easy to accept because of its direct nature of offering products to consumers. Telemarketing is usually used by several large companies to promote products they sell. Telemarketing (remote marketing) is the use of telephones and call centers to attract prospects, selling to existing customers, and providing services by taking services by receiving messages and answering inquiries over the phone as a new way of marketing using telecommunications technology as part of an organized and structured marketing program. (Saputra, 2017).

Telemarketing is also used by banks as a marketing strategy to offer products to consumers (Sulaehani, 2016). Banks are one of the most important roles in moving the economy of a country (Fauzi et al., 2020). Many banks using telemarketing services are considered to be able to increase bank income and achieve a target (Purbaya et al., 2020).

Telemarketing apart from helping to increase a company's revenue, other advantages can reduce sales costs, improve customer satisfaction, provide solutions for customers who have limited distance and time to be able to carry out transactions for the

protection program needed for both the customer himself and his family members (Dewi, 2016).

This is to determine the level of success of telemarketing in selling bank products, which can be done by applying and utilizing machine learning techniques (Fauzi et al., 2020). So this needs to be predicted to determine the success rate of bank telemarketing.

This research is not the first to be conducted, other researchers have also done the same thing being disgraceful based on the Table.1 Main Reference Paper is as follows :

Table 1. Main Reference Paper

Research Title	Method & Result Accuracy
A data-driven approach to predict the success of bank telemarketing	Comparing data mining models logistic regression, decision trees (DTs), neural network (NN), and support vector machine. Of the four models, neural network (NN) presents the best results (AUC = 0.8 and ALIFT = 0.7)
Feature selection with data balancing for prediction of bank telemarketing	Algorithm c4.5 with feature selection. Result ROC 95.60% TP Rate 92.14%

Source : (Masturoh et al., 2021)

Based on Table 1 Main Reference Paper, we can see that the paper entitled "A data-driven approach to predict the success of bank telemarketing" Researchers use 4 machine learning methods to predict the success of bank telemarketing and get the best results AUC 0.8 and ALIFT 0.7 by using the method neural network (NN) (Moro et al., 2014). Meanwhile, in the paper entitled "Feature selection with data balancing for prediction of bank telemarketing" researchers use the method c4.5 with feature selection with the resulting value ROC 95.60% TP Rate 92.14% Meanwhile, without using the data balancing technique and the selection of feature selection, the value is obtained ROC 88.40% TP Rate 91.19%, and if using a correlation-based feature subset algorithm without using the dataset balancing technique, the value is obtained ROC 91.00% TP Rate 91.26% (Vajiramedhin & Suebsing, 2014).

So in this study, we perform a classification to predict the success of bank telemarketing with data mining techniques using an algorithm's multilayer perceptron (MLP) by resampling. In this study, we used all attributes that is age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, days, previous, pot

outcome, and class y and preprocessing data by resampling because the data used is a data imbalance. method Multilayer perceptron (MLP) with resampling is done to get the best prediction results.

MATERIALS AND METHODS

In this study, we used a bank dataset taken from the website UCI repository on the link <https://archive.ics.uci.edu/ml/datasets/bank+marketing>. The dataset was published in 2012 with 17 features, namely age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, potoutcome, and y as the output variable. With class y as many as 45211 instances described in Table. 2 Details of Bank Telemarketing Dataset are as follows:

Table 2. Details of Bank Telemarketing Dataset

Features	Description	Type
1 age	client's age (years)	numeric
2 Job	Client's type of work	categorical
3 Marital	The client's marital status	categorical
4 Education	Client's highest education	categorical
5 Default	Does the client have credit?	categorical
6 Housing	Does the client have mortgage installments?	categorical
7 Loan	Does the client have a personal loan?	categorical
8 contact	Telephone contact that can be reached	categorical
9 Day	When was the last day you were contacted?	categorical
10 month	When was the last month contacted?	categorical
11 duration	How long did the last contact with the client	numeric
12 campaign	Number of contacts made during the promotion for this client	numeric
13 pdays	The number of days that pass after the client is contacted	numeric

Features	Description	Type
14 previous	The number of days that passed after the client is contacted	numeric
15 poutcome	The results of previous promotions	categorical
16 Emp.var.rate	Level of work variation	numeric
17 Cos.price.idx	Consumer price index	numeric
18 Cons.conf.idx	Consumer confidence index	numeric
19 Euribor3m	3-month Euribor rate	numeric
20 Nr.employed	Number of employees	numeric
21 label	Has the client subscribed to a time deposit?	categorical

Source: (Vajiramedhin & Suebsing, 2014)

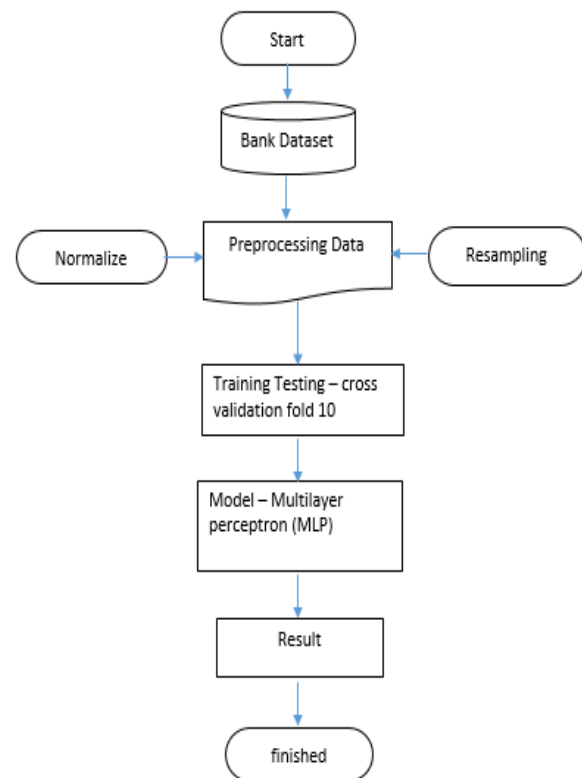
Experiments carried out on this dataset include using the multilayer perceptron algorithm classification method (MLP) (Purnama et al., 2014) which is an algorithm that adopts the workings of nerves in living things. The definition of the MLP architecture depends on the choice of the number of layers (layers), the number of hidden nodes (hidden nodes) on each layer, and the objective function of the approach. (Ramchoun et al., 2016) according to Teorema Kolmogorov, The MLP algorithm can approach various continuous functions with the weights and biases of the Single Hidden Layer network (SHL) which is optimized (Heidari et al., 2019). The Multilayer Perceptron method is part of the Artificial Neural Network. The Multilayer Perceptron method uses the Backpropagation neural network training algorithm (Pratiwi et al., 2019).

On the Multilayer Perceptron (MLP) The standard Sigmoid function is used where the weighted amount of several inputs can be inputted to the activation level via the transfer function to produce the output, and the units are arranged in a feed-forward topology layer called a Forward Network feed. When there is more than one hidden layer, the output from the hidden layer is entered into the next hidden layer and a separate weight is used for the summation to each subsequent layer (Amrin & Satriadi, 2018).

But before classification, first, preprocessing normalization is done. Use Normalization to process the scaling of attribute values from data so that they can lie within a certain range (Nasution et al., 2019). Many data have different ranges so that normalization must be carried out.

In addition to normalization techniques on preprocessing, Resample technique was also performed in this study, This method works by reducing the population of the majority class in such a way that it is equal in number to the minority class (Pujiyanto, 2016). This reduction can be done randomly in this case called random undersampling or it can be done using some statistics in this case called information undersampling. Some of the information on the undersampling method and the iteration method also applies the data cleaning method to further refine the majority class sample (Shelke et al., 2017)

The stages of the research methodology carried out are described in Figure 1 Research Methodology is described below :



Source : (Masturoh et al., 2021)

Figure 1 Research Methodology

In Figure 1, we can see how the stages of the research were carried out, the dataset was taken from the website UCI repository then performed data preprocessing by normalizing and resampling, then the data is in training and testing with cross-validation fold 10. After the preprocessing is carried out, the classification is carried out using a machine learning algorithm, namely the Multilayer Perceptron (MLP), After that, the results of the correlation pattern between data will be obtained that can predict the success of bank telemarketing data as seen from the accuracy, TP Rate, PCR Area,

ROC Area, if it continues it can be implemented into a prediction software.

The platform used in this study has specifications based on table 3. The research platform is as follows:

Table 3 Platform Penelitian

Processor	AMD E2-2000 APU with Radeon(TM) HD Graphics 1.75 GHz
System type	32-bit Operating System
RAM	4,00 GB
Software	Aplikasi WEKA 3.8

Source : (Masturoh et al., 2021)

RESULT AND DISCUSSION

The results of the research methodology carried out 2 comparisons without using resampling and using data resampling. The results obtained are described in the following discussion table :

Table 4 Test Results

Method	TP Rate	FP Rate	ROC Area	Accuracy
Without resample	0.902	0.602	0.890	90.179%
With resample	0.943	0.260	0.899	94.271%

Source : (Masturoh et al., 2021)

In table 4, it can be seen that the test results on the bank dataset using the multilayer perceptron algorithm (MLP) are tested without a resample, the results are obtained TP Rate 0.902, FP 0.602, ROC Area 0.890, and accuracy value 90.179%. Meanwhile, using the resample, the results obtained TP Rate 0.943, FP Rate 0.260, ROC Area 0.899, and 94.271% accuracy value. The biggest accuracy result can be seen that the prediction of the success of bank telemarketing data is using the multilayer perceptron algorithm (MLP), namely by using resampling.

CONCLUSION

Telemarketing is considered effective for increasing a company's income. However, to determine whether bank telemarketing is effective or not in increasing a profit in a company, data mining techniques are needed to predict the success of telemarketing, including from bank telemarketing. In this study, the proposed method is to use a multilayer perceptron (MPL) algorithm with resampling on a bank dataset consisting of 17 attributes with 1 class proven to get the highest

accuracy value by using resampling of 94,271%, TP Rate 0.943, FP Rate 0.260, ROC Area 0.899 and when compared to without using resampling. The purpose of using resampling is a class whose data imbalance in the dataset used can be overcome with sample sampling techniques so that the majority of data is discarded by replacing it or without replacing it. Further research that can be carried out by subsequent researchers can implement the formed patterns and build software to predict the success of bank telemarketing.

REFERENCES

- Amrin, A., & Satriadi, I. (2018). Implementasi Jaringan Syaraf Tiruan Dengan Multilayer Perceptron Untuk Analisa Pemberian Kredit. *Jurnal Riset Komputer (JURIKOM)*, 5(6), 605–610.
- Dewi, S. (2016). Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *None*, 13(1), 60–66.
- Fauzi, A., Wati, F. F., Sulistyowati, I., Akbar, M. F., Rahmawati, E., & Sari, R. K. (2020). Penerapan Metode Machine Learning Dalam Memprediksi Keberhasilan Panggilan Telemarketing Menjual Produk Bank. 6(2), 213–222.
- Heidari, A. A., Faris, H., Aljarah, I., & Mirjalili, S. (2019). An efficient hybrid multilayer perceptron neural network with grasshopper optimization. *Soft Computing*, 23(17), 7941–7958. <https://doi.org/10.1007/s00500-018-3424-2>
- Masturoh, S., Nugraha, F. S., Nurlela, S., Saelan, M. R., & Saputri, D. U. E. (2021). *PREDIKSI KEBERHASILAN TELEMARTETING BANK MENGGUNAKAN ALGORITMA MULTILAYER PERCEPTRON (MLP) DENGAN RESAMPLING*.
- Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. <https://doi.org/10.1016/j.dss.2014.03.001>
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Pratiwi, P. G., Putra, I. K. G. D., & Putri, D. P. S. (2019). Peramalan Jumlah Tersangka Penyalahgunaan Narkoba Menggunakan Metode Multilayer Perceptron. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 7(2), 143. <https://doi.org/10.24843/jim.2019.v07.i02.p>

06

- Pujianto, U. (2016). Strategi Resampling Berbasis Centroid Untuk Menangani Lunak. *Teknno*, 25(Maret), 1-6.
- Purbaya, M. E., Nugraha, A. F., Gustina, S., & Azis, M. K. (2020). Meta-Algorithms untuk Meningkatkan Kinerja Klasifikasi dalam Keberhasilan Telemarketing Perbankan. *Techno.Com*, 19(4), 385-396. <https://doi.org/10.33633/tc.v19i4.3725>
- Purnama, N., Putra, I. K. G. D., & Bayupati, P. A. (2014). Klasifikasi Website Menggunakan Algoritma Multilayer Perceptron. *Teknologi Elektro*, 13(2), 9-15.
- Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y., & Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(1), 26. <https://doi.org/10.9781/ijimai.2016.415>
- Saputra, E. P. (2017). Prediksi Keberhasilan Telemarketing Bank Untuk. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 2(2), 66-72.
- Shelke, M. S., Deshmukh, P. R., & Shandilya, P. V. K. (2017). A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique. *International Journal of Recent Trends in Engineering and Research*, 3(4), 444-449. <https://doi.org/10.23883/ijrter.2017.3168.0uwxm>
- Sulaehani, R. (2016). Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination. *ILKOM Jurnal Ilmiah*, 8(3), 182-189. <https://doi.org/10.33096/ilkom.v8i3.83.182-189>
- Vajiramedhin, C., & Suebsing, A. (2014). Feature selection with data balancing for prediction of bank telemarketing. *Applied Mathematical Sciences*, 8(113-116), 5667-5672. <https://doi.org/10.12988/ams.2014.47222>

