Jurnal Ilmiah PPI-UKM
Scientific Journal of PPI-UKM
Persatuan Pelajar Indonesia
Universiti Kebangsaan Malaysia
Selangor, Malaysia
www.ppiukm.org

Scientific Jounal of PPI-UKM

Science and Engineering

Vol 7, No 2 (2020)
ISSN No. 2356 - 2536

Classification Data for Direct Marketing using Deep Learning

Amril Mutoi Siregar[a]*, Sutan Faisal[a], Hanny Hikmayanti Handayani[a], Asep Jalaludin[b]

[a]*Buana Perjuangan University, Jl. Ronggowaluyo, Karawang 41311, Indonesia*
[b]*MIC STMIK, Ruko Harco Teknik, Cikarang Bekasi 17550, Indonesia*

Abstract

One of the tasks of banking marketing is to analyze customers' data and to find out the potential customers to save deposits. Generally, the method used to analyze customer data is by classifying all customers who have taken the time deposit into the target marketing, so this method causes the high cost of marketing operations. Therefore, this research is conducted to help solve the problem by designing a data mining application that can serve to classify the criteria of customers who potentially to save deposits in the bank. In classifying customer data has been done a lot by researchers before with various algorithms, now researchers use deep learning to classify the target in want by the banking. The results showed that achieved using deep learning accuracy is = 80%, MSE = 0.0943, AUC = 0.8533. The results of this study can be reference to build an application that can facilitate the banking in obtaining its target marketing in the future.

Keywords: Datamining, Deep learning, Artificial Intelligency, Backpropagation, Classification.

Received: 8 April 2018; Accepted: 24 March 2019

1. Introduction

In the banking world often found the problem about how to find customers and / or raise funds, either deposits, deposits or credits. It can not be denied that many problems are found in banking, especially credit taking, for example bad credit. The largest banking revenue is from the credit sector, so the credit sector is considered the most strategic for growing banking.

The method used to determine the target market so far is to classify all customers who have paid their credit installment into the next marketing target, but not necessarily all the customers will be interested in borrowing the loan back. Therefore, to increase the loyalty of bank customers, this section is required to be able to take the right decision in determining the marketing strategy, both products and banking services that will be offered to customers, especially in credit problems. This can be realized if we obtained enough information to assist Decision Making Process in the field of marketing, one of them is to classify or predict customer criteria for the purposes of marketing targets. Data mining methods are an appropriate step for marketers to analyze the market so they can understand existing markets or find new opportunities to increase profits including predicting target market criteria. With the case above this research focus on data mining utilization to predict and classify the criteria of credit customers, so it can be known whether the customer is a potential customer of credit. Datamining method used in this research is deep learning and the data used are direct marketing datasets at Bank. The dataset is often conducted for research use in the field of datamining.

2. Data Mining

Datamining is process of finding new relationships that have meaning, patterns with most of data stored in storage device using technology, pattern recognition is statistical technique and mathematical.

Datamining is combination of several disciplines that combine techniques in machine learning, pattern recognition, database statistics, and visualization to handling information retrieval problem from every large database [1].

Data mining is analysis of datamining (bigdata) to find a clear relationship and summarize it that has not been known before with the latest ways of understanding and useful for the owner of the data [2]. Datamining is process that uses statistical techniques, Mathematics, artificial intelligence, machine learning to extract and identify useful information related to large databases. And datamining is series of process to explore the added value of dataset from of knowledge that has not been known manually [2]. And datamining can be interpreted as follows:

1. Datamining is an automated process for analyzing data.
2. The data to be processed is very large data, those it is difficult to process manually.
3. The purpose of datamining is to get a relationship or pattern / technique that gives a useful indication.

Data mining is nothing new, one of difficulties to define datamining is fact that Datamining inherits many aspects

* Corresponding author. Tel.: +62 813-1639-3054
E-mail address: amrilmutoi@ubpkarawang.ac.id
DOI: 10.27512/sjppi-ukm/se/a15052020

and techniques from previously established fields of science. datamining has long roots from different fields of science such as artificial intelligence, machine learning, statistics, databases, and also information retrieval [2].

Databases stored in storage media are rarely used by most users and even within a certain time the data is removed because it is considered waste and only meet the storage media only. The assumption is not entirely true, because in fact large databases can provide the information needed for various purposes, both for business interests in decision making and for science and research.

3. Classification

One of task that can be done with data mining is classification. The classification was first applied to the plan of plants that classified a particular species, as conducted by Carolus Linnaeus who first classified species based on physical characteristics. Furthermore, he is known as the father of classification [3]. In the classification there are target category variables. The methods/models developed by researchers to solve classification cases include Decision Tree, Artificial Neural Networks. Etc. In this research, the concentration of the workings of the deep learning toward the classification of marketing data bank. So, the expected results of the deep learning performance. H2O Deep Learning Operators are used to predict Survived attributes from data marketing collection. The label is binominal, the classification will be done to analyze the quality of the model, and the Split Validation operator is used to generate training and testing datasets. Deep Learning operator parameters are the default values, meaning 2 hidden layers, each with 50 neurons built. Example set which has label is connected to the Performance operator (Binominal Classification), which calculates the value of the Accuracy metric. In the Deep Learning Model output process, the data labeled and Performance Vector are generated.

This operator is used to evaluate the statistical performance of binominal classification tasks, classification assignments in which label attributes have binominal types. This operator lists the performance criteria values of binominal classification task.

Operators are used specifically to evaluate performance in the binominal classification which is a classification where the label attribute has a binominal type. Many other performance evaluation operators are also available on RapidMiner for example. Performance operator (Classification), Performance operator (Regression), and others. The Performance Operator (Binominal Classification) is used for binominal classification. On the other hand, the Performance operator automatically determines the type of learning task and calculates the most common criteria for that type. we can use the Performance (User-Based) operator if we want to write their own performance measures.

4. Deep Learning

Deep learning is a stack or a number of several algorithms or methods, thereby evolving into a deep learning approach with various architectures. Some of the objectives of this method are feature extraction, to take advantage of all the resources as optimally as possible. Since most of the data in the world is not labeled (uncategorized), deep learning is usually a stack of unsupervised and supervised learning stack algorithms that can utilize labeled or unlabeled data. With the utilization of more optimal information will certainly improve the performance of the resulting model. The most common approaches to implement deep learning are graphical methods or multilayer representations, or multilayer graphical models such as belief network, neural network, and others. Basically, methods as ordinary machine learning is a statistical and stochastic method that is already widely known in the world of mathematics, especially statistics. In other word the information shown in the hidden layer or the value of each unit on hidden layer can represent input layer or the extract from information available on input layer, and the corresponding synapse of weight is transferred to the network on the deep neural network, as the initialization between the synapse input layer and the first hidden layer. Next execute autoencoder back between hidden layer 1 and hidden screen 2, and so on until hidden layer to n-1 and to n. Perhaps technically simple as Deep Network, and actually this is a fairly simple algorithm. There are several other algorithms like deep autoencoder, Boltzmann Machine, stacked, deep belief network and others. Deep learning itself is often paired with bigdata, which is analyzing bigdata with its 3V by using deep learning. So, looking at the complexity of these methods required supporting technologies such as hardware and parallel or cluster implementation approach either from the implementation of algorithms or data management quite fast, so as to suppress the complexity of time that appears in deep learning. There are various approaches done from starting parallel programming using the usual multi core, GPU, to message passing interface for example utilizing more than one machine. As shown in google, Microsoft etc.

4.1 Artificial Neural network (ANN)

ANN is inspired by the awareness of the complex learning system in the brain that consists of sets of neurons that are closely related. Neural networks are capable of performing very complex tasks such as classification and understanding patterns. ANN can estimate a wide range of statistical models and be flexible in describing models (linear and on linier) [5]. ANN can be used for the same problems with multivariate statistical problems such as multiple regression, discriminant analysis, and cluster analysis. In many cases, the results obtained with ANN can be compared with the multivariate statistical model**.**

ISSN No. 2356 - 2536
DOI: 10.27512/sjppi-ukm/se/a15052020

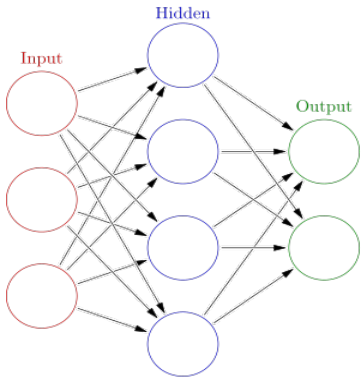
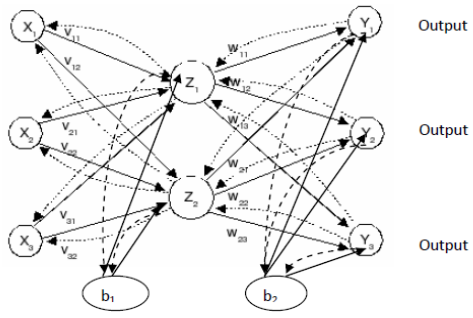


Figure 1. ANN Network Architecture

In the ANN network architecture, there are 3 layers of the input layer as the input value to be trained, then the value of the input is sent to the hidden layer by counting it with the weights, then from hidden layer sent again to output layer in same way is by calculating the weight - the weight of existing. The output layer contains the value of the target that will be the reference in the testing phase of the model (testing). From the results of this training will get a model of trained Neural Network where the value of its weight - weight that will determine the results of the prediction or classification when testing phase.

4.2 Backpropagation

Neural Network (NN) is a collection of elements of interconnected simple process elements is called neuron. Each Neuron is associated with another neuron with a direct communication link through pattern relationships is called Network Architecture [6]. Each of these relationships has a weighted connection (weight) that is trained to achieve response. Those by training of the data based on the connection weights are expected to get output. The method used to determine the connection weight is called training algorithm (training algorithm). Backpropagation has several units in one or more hidden screens.



Description: Y1…Y3: Output, X1...X3: input, b1, b2: bias, Z1, Z2: hidden layer

Figure 2. Backpropagation Architecture (Fausett, 1994)

Figure 2 is the backpropagation architecture with n as input plus bias, a hidden layer consisting of p plus bias, and m is the output. In backpropagation training uses the minimum point search method to look for weights with minimum errors. The backpropagation algorithm uses an error output to change its weight value backwards [5].

Backpropagation training consist 3 phases: advanced phase, reverse phase and weight change phase. The first phase is feedforward step, at this step the initial process for initialization of input weights is done. This step it also determined the learning rate (α), the value of error tolerance and the epoch (cycle of each training pattern) required during the computation process takes place. After all the initialization process is done, the next step is the process forward. Each input unit Xi will send input signal to hidden layer. Once calculated using the activation function then the output will be sent to the layer above it, that is the output layer. After the output value is obtained, it is compared with the actual output target. The difference in output value – output target is called error. If the error value is less than or equal to the threshold value, the shrimp process is stopped, but if not the error value is used to modify the weights to correct the error. The second step is backward or backpropagation. This step the value (δk) obtained on the output layer is used to correct the weights in the hidden layer directly related to the output layer. After the error value (δj) in each unit in the hidden layer is also calculated to correct the weights that connect the input layer with hidden layer. The third step is the correction to correct weights step. After all the weights in the input layer and hidden layer are modified according to the error factor, the three stage are repeated continuously until the stop condition. The stop condition in question is if the specified epoch number is reached or if the error value of the network is equal to or less than the predefined error tolerance value. In the training step, the network is expected to train all the training data provided to obtain the final weight of the network to be used during the testing step.

5 Methodology

In this study using rapidminer studio 7.4 testing tools, using methodology as follows:

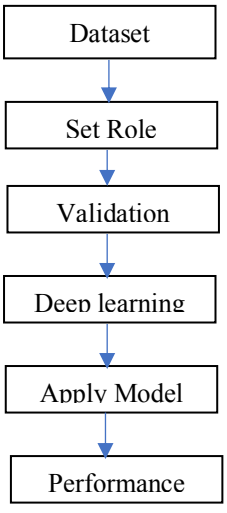


Figure 3. Methodology in use

5.1 Dataset

Dataset is collection of data that will be the object of research. dataset is a representation of a database, data matrix that has columns to represent of attributes and rows to represent of the amount data. in this research using datasets about bank customers.

5.2 Set role

Set role is a stage to determining attributes for other operators as targets. because the dataset has many attributes, this set role stage determines the label of the purpose of this study.

5.3 Validation

Validation as operator to get simple validation is to randomly divide the dataset into a set of training, determine the test and evaluate the model. This operator performs split validation to estimate the performance of the learning operator, mainly used to estimate the accuracy of a model that is displayed as a result.

5.4 Deep learning

Deep learning is an algorithm that uses to determine the expected results. As for the type of deep learning used multilayer feed forward networks, trained with the slope of the stochastic gradient using backpropagation.

Deep learning is used to classify and predict the bank customers dataset. Label is used binominal, classification will be performed. The correct the quality of the model, split validation operator is used to generate the training and testing datasets. Deep learning operator parameter are the default values. This means that 2 hidden layers, each with 50 neurons will be constructed. Labelled dataset is connected to be performance binominal classification operator, calculates the accuracy metric. On the process output the deep learning model, labelled data and performance vector is shown.

5.5 Apply Model

Apply model is a model that is trained in datasets by other operators, often as learning algorithm. After that, this model can be applied to other datasets, the goal is to get a classification or prediction on the data that is not visible to transform data with the model process. The dataset used must be compatible with the attributes of the model, meaning the dataset has the same number, type, sequence, and role attributes as the dataset used to produce the model.

5.6 Performance

Performance is used as an instrument tools for statistical perform evaluation in the classification of two values. This classification for assignments in which label attributes have binominal types. This operator lists the performance criteria values of binominal classification tasks.

To measure the results of this study by using confusion matrix, classification recall, Classification prediction, accuracy.

Accuracy is the percentage of number data records classified or predicted correctly by the algorithm.
Accuracy Formula:

$$Accuracy = \frac{True\ Yes + True\ No}{Number\ data} \quad (1)$$

Misclassification error rate is the percentage of the number data records for classified and predicted false.
Misclassification Error Rate Formula:

$$Misclassification\ Rate = \frac{False\ Yes + False\ No}{Number\ data} \quad (2)$$

6. Result And Discussion

6.1 Results of Data Set Analysis

The dataset used is the set of marketing data Bank taken from the website [7], this data set contains some data information about customers who owned bank. The total data on this dataset amounted to 200 records which each record has 17 attributes including:

1. Age (Numeric)
2. Job (Polynomial)
3. Marital (Polynomial)
4. Education (Polynomial)
5. Default: has credit in default (Binominal)
6. Balance (Numeric)
7. Housing (Binominal)
8. Loan (Binominal)
9. Contact (Polynomial)
10. Day (Numeric)
11. Month (Polynomial)
12. Duration (Numeric)
13. Campaign (Numeric)
14. P-days (Numeric)
15. Previous (Numeric)
16. P-outcome (Polynomial)
17. Y: Target (Binominal)

From the attribute Dataset above (1 to 17) will be the training & test process using deep learning method, while the 17th attribute will be the target result of the classification process. and in this paper the writer will try to analyze the difference of accuracy and error obtained by making changes to the number neurons in hidden layer.

6.2 Experiment Result and Evaluation

In this experiment there are 17 attributes that will be in training and 2 values indicating the target (classification) on the 17th attribute which means in deep learning in the

ISSN No. 2356 - 2536
DOI: 10.27512/sjppi-ukm/se/a15052020

initialization of 16 neurons on the input layer, 2 neurons in the output layer.

The results of this study are:

Table 1
Performance deep learning.

Measurement	Result
Accuracy	80 %
Class Precision (No)	91.49 %
Class Precision (Yes)	38.46 %
Class Recall True No	84.31%
Class Recall True Yes	55.56%

Here's the performance of deep learning with binominal matrix model in mining: MSE, R2, AUC result as below:

DeepLearning

```
Model Metrics Type: Binomial
 Description: Metrics reported on full training frame
 model id: rm-h2o-model-deep_learning-629690
 frame id: rm-h2o-frame-deep_learning-300493
 MSE: 0.09431424
 R^2: 0.2602805
 AUC: 0.85333335
 logloss: 0.3109382
 CM: Confusion Matrix (vertical: actual; across: predicted):
        no  yes   Error        Rate
    no  160   10  0.0588  = 10 / 170
   yes   13   17  0.4333  =  13 / 30
Totals  173   27  0.1150  = 23 / 200
Gains/Lift Table (Avg response rate: 15.00 %):
```

Figure 4. Result of Deep Learning Models Matrix

7 Summary

The method of classification by using deep learning is very good for determining the truth of the classification in data mining. Proven with the result of accuracy = 80%, AUC = 0.8533, MSE = 0,0943. The use of deep learning in datamining is still limited binominal**.**

8 Future work

After successfully performing this research with good results it is necessary to do learning machine using deep learning on datamining both classification and prediction in order to be able to convince deep learning performance runs well.

9 Reference

[1] Larose, et al, 2005, Discovering Knowledge in Data Mining an Introduction to Data Mining, Wiley inter science.

[2] Bramer, et al, 2007, Principles of Data Mining, Springer Science.

[3] Mardi, et al, 2014, Analisa Data Rekam Medis Untuk Menentukan Penyakit Terbanyak berdasarkan international classification of Disease (ICD) Menggunakan Decision Tree. UPI YPTK Padang.

[4] Bengio Y, et al, 2013. Representation Learning A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Learning. 35(8): 1978-1828.

[5] Nielsen. M, 2016. Neural Networks and Deep Learning. http:/neuralnetworksanddeeplearning.com.

[6] Fausset, et al, 1994. Fundamentals of Neural Networks. New Jersey: Prentice Hall Inc.

[7] Moro, et al, 2011. Using Data Mining for bank Direct Marketing. An Application of the CRISP-DM Methodology. In P. Novais, Proceedings of the European Simulation and Modeling Conference-ESM2011, pp.117-121, Guimaraes, Portugal, EUROSIS.

ISSN No. 2356 - 2536
DOI: 10.27512/sjppi-ukm/se/a15052020