



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 4 - V7I4-1484)

Available online at: <https://www.ijariit.com>

Analysis of bank telemarketing data using Microsoft Azure machine learning

Jaswanth Kankanala
jaswanth.r.kankanala@outlook.com
MODAK, Analytics, Gachibowli, Telangana

Sravya Sri V.
Sravysrivitalam@gmail.com
Caterpillar

Venkatesh narahari
Naraharivenkatesh@outlook.com
KPMG

ABSTRACT

Nowadays, marketing expenditure in financial business is enormous, implying that it is fundamental for banks to upgrade promoting procedures and further develop adequacy. Understanding clients' need prompts more viable advertising plans, more brilliant item plans and more prominent consumer loyalty. The aim is to predict customer response for future need marketing campaigns to increase the effectiveness of bank telemarketing campaign. Here in this paper, we will empower the bank to foster a more granular comprehension of its client base, foresee clients' reaction to its selling crusade and build up an objective client profile for future advertising plans. By investigating client highlights, like socioeconomics and exchange history, the bank will want to anticipate client saving practices and recognize which clients are bound to set aside term installments. The bank would then be able to zero in its promoting endeavors on those clients. This won't just permit the bank to get stores all the more successfully yet, in addition, increment consumer loyalty by decreasing unwanted ads for specific clients.

Keywords: Bank Marketing, Telemarketing, Machine Learning, Logistic Regression

DATA

From May 2008 to November 2010, a Portuguese financial organization used direct phone marketing to sell deposits among current clients. It is freely accessible through the UCI Machine Learning Repository. There are 41,188 observations with no missing values.

- The dataset captures 16 input variables for each observation, including qualitative and quantitative consumer factors such as age, employment, housing and personal loan status, account balance, and several interactions.
- A single binary output variable denotes "yes" or "no," revealing the outcomes of the marketing phone calls. "Yes" means that a customer subscribed to term deposits.

ATTRIBUTE AND THEIR INFORMATION

Attribute	Attributes information
Age	Age of person
Job	Type of job
Marital	Marital status
Education	Educational details
Default	Has credit in default
Housing	Has housing loan?
Loan	Has personal loan?
Contact	Contact communication type
Month	Last contact month of the year
Day of week	Last contacted day of the week
Duration	Last contact of duration, in seconds
Campaign	Number of contacts performed during this campaign
Pdays	The number of days that passed when the client was last contacted.
Previous	Number of contacts performed before this campaign
Poutcome	The outcome of the previous marketing campaign
Emp.var.rate	Employment variation rate

Cons.price.idx	Consumer price index
Cons.conf.idx	Consumer confidence index
Euribor3m	Euribor 3 month rate
Nr.employed	Number of employees
Y (output variable)	Has the client subscribed to a term deposit

LOGISTIC REGRESSION

Logistic regression is a statistical technique for evaluating a dataset in which one or more independent factors influence the result. A dichotomous variable is used to assess the outcome (in which there are just two possible results.)

Creating a Logistic Regression Model on Microsoft Azure Cloud(Machine Learning): We construct a logistic model that can predict two (and only two) outcomes using the **Two-class logistic regression** model. In our experiment, we are using 2-class classification model, which is logistic regression, which gives the outcome as "yes" or "no" and 0 or 1 and "true" or "false." Logistic regression is a statistical method used to predict the likelihood of an outcome and is exclusively common for all classification tasks. The algorithm predicts the probability of the occurrence of the event for fitting data to a logistic function.

CONFIGURING TWO-CLASS LOGISTIC REGRESSION IN AZURE

(Machine Learning): To train the model, we must give a dataset with a class column or label; this model is exclusively meant for two-class issues with labels that have precisely two values. For example, the label variable's potential values are "Yes" or "No," "High" or "Low," "Win" or "Loss," "1" or "0."

Implementation Steps - logistic regression

- Drag and drop the Data set (Bank marketing data.csv) to the canvas of AML.
- Click output circle to visualize
- Drag-and-drop 'select column from the dataset' to select the column.
- Find 'Logistic Regression' and drag it onto the canvas
- Click on 'Logistic Regression' and make sure that the 'Ordinary Least Squares' solution technique is selected in the properties window
- Look for 'Train Model,' then drag & drop it into the canvas
- Connect the output of 'Logistic Regression' to the left input of the 'Train Model' and the output of select column from the dataset' to the right input of the 'Train Model.'
- In the properties box, click on 'Train Model,' then pick the launch column selector.
- Select the column(Creditability) for which the prediction has to be done
- Drag-and-drop 'Score Model' from the left pane and uncheck the 'Append score column' in the properties window
- Connect the 'Train Model' output to the left input of the 'Score Model' 'select column from the dataset' to the right input of the 'Score Model.'
- Drag-and-drop 'Evaluate Model' from the left pane
- Connect the output of 'Score Model' to the input of 'Evaluate Model'
- Click on Run
- After execution, click on the output circles of 'Train Model,' 'Score Model' and 'Evaluate Model' to see the value of R- squared
- In the experiment click again click on "set up web service" in the bottom pane
- Select Retraining Web Service
- Click on "RUN" to run in the bottom pane
- After execution, again click on "set up web service" in the bottom pane and select Predictive Web Service
- Again, click on "RUN" to run in the bottom pane
- After execution, click on "Deploy web service" it will deploy and take you to the web service page
- Click on the Test button, enter data to be predicted
- Predict window will open with the output

Logistic Regression: Training Experiment



Figure 1: Training Experiment for Bank Tele Marketing

(Logistic Regression)

Two-Class Logistic Regression

Create trainer mode

Single Parameter

Optimization tolerance: 1E-07

L1 regularization weight: 1

L2 regularization weight: 1

Memory size for L-BFGS: 20

Random number seed

Allow unknown categ...

START TIME	7/22/2021 ...
END TIME	7/22/2021 ...
ELAPSED TIME	0:00:00.000
STATUS CODE	Finished
STATUS DETAILS	Task output was present in output cache

Figure1.1: (Logistic Regression) Parameters

Logistic Regression: Predictive Experiment:

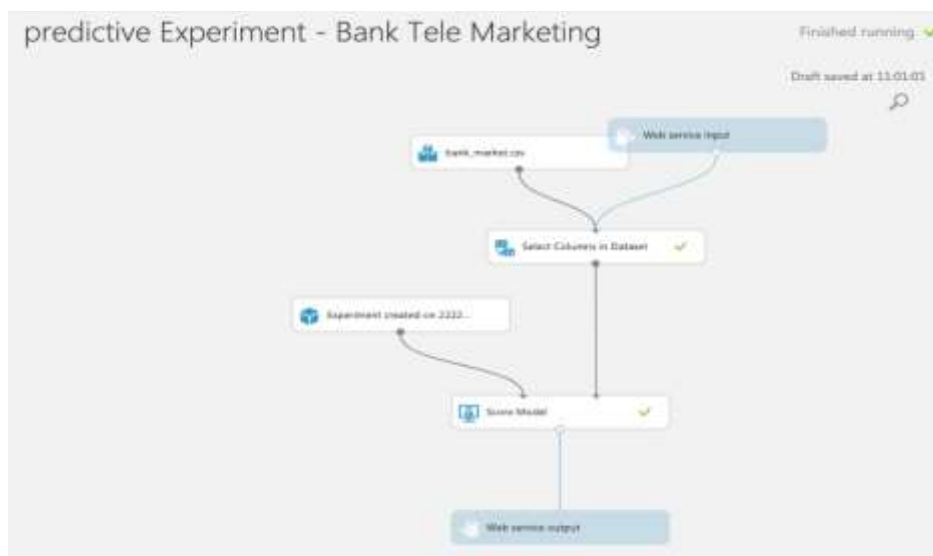


Figure 2: Predictive Experiment for German Credit Data

(Logistic Regression)

In AML, what does a trained model do?

The process of training a classification model is an example of supervised machine learning. That is, you should offer a dataset containing past data from which to derive patterns. The data must include both the outcome we are attempting to predict and any associated factors (variables). The machine learning model uses the data to retrieve statistical patterns and create a model.

How do you train a model?

- Include the Train Model module in the experiment. This module may be found in Azure Machine Learning Studio under the Machine Learning category. Drag the Train Model module to your experiment by clicking on Train.
- Attach one of the categorization models provided by Azure Machine Learning Studio to the left input.
- Attach training data to the Train Model's right-hand input.
- For the Label column, you should pick a single column that includes model results that we may utilize for training. Now, select the column in the dataset that contains the values you wish to forecast by clicking the Launch column selector. The label column should have categorical or discrete values in classification scenarios.
- Run the experiment. If we rendered more data, then it takes more time than usual.

Score model?

- We can utilize the score model to generate predictions using a trained classification model; projected values can be in several formats varying on the model and the input we provide.
- If we use regression models, the score model produces a projected value for the class and the probability of the anticipated value.
- In the case of regression models score model generates just predicted numeric value.
- After generating the set of scores by using score models, we can connect scored dataset to evaluate a model for finding metrics and performance of the model.

The output of logistic regression for German credit:

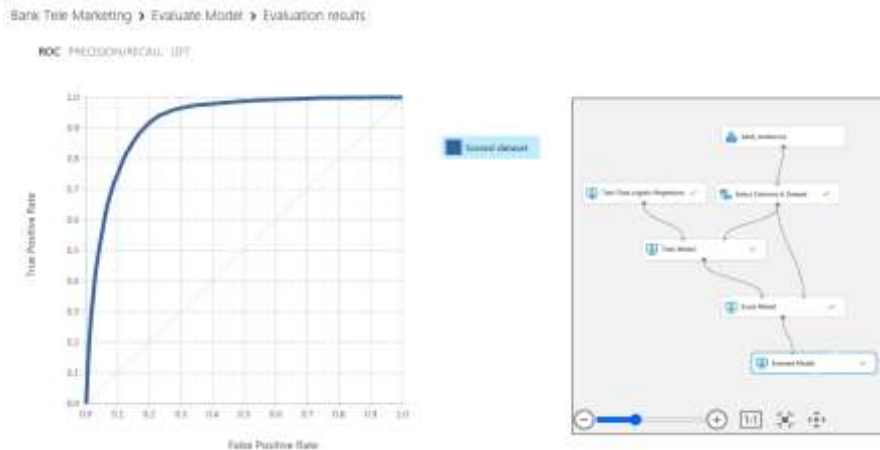


Figure 3: Roc curve

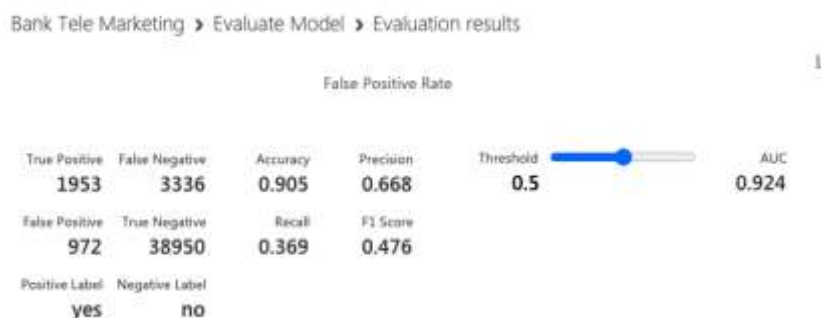


Figure 4: Confusion Matrix & Accuracy when the threshold is 0.5

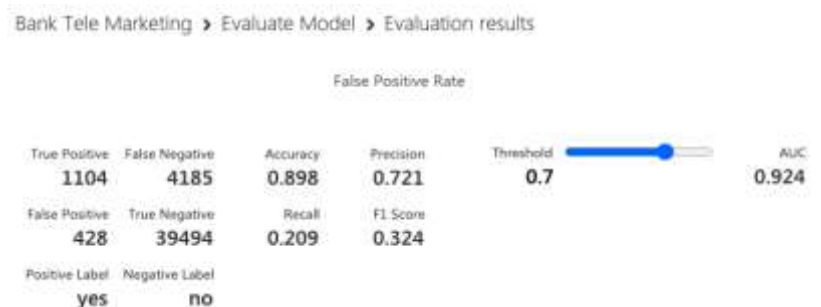


Figure 5: Confusion Matrix & Accuracy when the threshold is 0.7

REFERENCES

- [1] Barboza, F., Kimura, H. & Altman, E. (2017), 'Machine learning models and bankruptcy prediction, Expert Systems with Applications 83, 405–417.
- [2] Bardsley, W. E., Vetrova, V. & Liu, S. (2015), 'Toward creating simpler hydrological models: A lasso subset selection approach,' Environmental Modelling & Software 72, 33–43.
- [3] Beucher, A., Møller, A. & Greve, M. (2017), 'Artificial neural networks and decision tree classification for predicting soil drainage classes in denmark', Geoderma .
- [4] Cramer, S., Kampouridis, M., Freitas, A. A. & Alexandridis, A. K. (2017), 'An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives', Expert Systems with Applications 85, 169–181.
- [5] Friedman, J., Hastie, T. & Tibshirani, R. (2001), The elements of statistical learning, Vol. 1, Springer series in statistics New York.
- [6] Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. (2010), 'Variable selection using random forests', Pattern Recognition Letters 31(14), 2225–2236.
- [7] Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. (2015), 'Vsurf: An r package for variable selection using random forests.', R Journal 7(2).
- [8] Gil, D. & Johnsson, M. (2010), 'Using support vector machines in diagnoses of urological dysfunctions', Expert Systems with Applications 37(6), 4713–4718. Izetta, J., Verdes, P. F. & Granitto, P. M. (2017), 'Improved multiclass feature selection via list combination', Expert Systems with Applications 88, 205–216.