# **Predicting Success of Bank Direct Marketing By OSAS**

ID:773950464

NAME: Yong Wan

UPI: wany991

## 1.Business and Situation understanding.

“First is developing an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the customer’s viewpoint.” (Fayyad et al., 1996)

Banks are a data-intensive industry and data is a crucial factor of competition, whether it is an offline business or online business. For the banking industry, it is not the most critical to choose the platform or technology, the key factor is to understand the demand of business. Mining the rules in the data banks had is a key competitive ability in the market. Citibank has been a good example. It has signed an agreement with IBM to explore the potential uses of IBM Watson in 2012. IBM Watson was able to be used to process a large amount of customer data and analyze customer needs.

It Has made people have no difference in distance in communication by the rapid development of the technology. It also enables consumers to change their roles e easier, from passive roles in tradition to active creators. Therefore, how to find new profit growing points, needs of customers and comprehensive personalized services is a big problem all bank face today. It is lucky that the development of information technology will be able to convert passive customer service to proactive customer care. telemarketing has become a very important marketing event.

Telemarketing first appeared in the United States in 1970s. It grew quickly With the popularity of mobile phones. Enterprises have begun to adopt this new marketing method, constantly expanding the size of customers, tapping the potential value of customers, forming a customer-centered market.Generally speaking, centralized customers are managed by call center by remote operation. The center communicates with customers through various channels. In the family, telephone (mobile phone or fixed telephone) is the most widely used. We would think that telemarketing means marketers are selling through Hawking. In fact, the definition of telemarketing will be more extensive. According to the difference between sender and receiver, telemarketing can be divided into two categories: exhaled telephone and access telephone.

1. call out

The reason for outgoing calls is active marketing because the originator is a company. Main business of exhaling business Including: Market Research and commodity sales.

1. call in

The founder of marketing is customer, so telephone is also known as response to telemarketing. The boss of the telephone company. Including: air ticket booking and dismantling.

Telephone as a marketing channel has its own advantages. Of course, there are also some shortcomings.

The advantages of marketing include:

1. providing instant feedback from customers is conducive to the interaction of marketers in marketing activities.
2. this is a fruitful marketing channel.
3. it is a relatively sensitive marketing channel. Through the timely feedback from customers, telephone salesmen can use electricity.
4. Adjust and change the content of the word accordingly.

The main disadvantages of telemarketing are as follows:

1. respondents often feel insulted or offended. Because telemarketing usually does not require prior approval.

The consent of the interviewer will make the client feel offended and may have an impact on marketing activities.

1. you can't see each other. You can't see each other's performance except voice.

The data set from Portugal has 41188 telephone contact records. The goal is to sell time deposits through telephone. So it is the main task to predict whether customers subscribe to time deposits (variable y). According to the characteristics of the data set and the goal to be achieved, the data is pre-processed and different DM models are selected for data mining. Then the model is evaluated, and the test set is used to forecast, providing bank telephone marketing related suggestions. Choose the best prediction model for landing application.And put forward some suggestions for customer service center telephone marketing.

### Base on the data, there were only 10% of success in the past three years, it means there is a massive gap between success and unsuccess. Also, we should do something to improve it to decrease the phone call times due to our customers may be annoyed and increase the profit for banks due to the phone agents get more success by same phone outs as before. So we aim to decrease 40% of the total amount of phoning out and at max decrease 5% of the total of successful calls. This means we will get at least 20% of success of phoning ,double than before. Also, it means we should get at least 90% correct prediction on customers who's original answer is Yes and nearly 40% accurate prediction on the customers who's original answer is No.

### This objective is to predict the success of phoning a customer then decided whether he was worth to call.

### In conclusion, the main tasks includes:

### 1.Describe the methods for model assessment (for example, accuracy, performance, etc.).

### 2.Define benchmarks for evaluating success. Provide specific numbers.

### 3.Define subjective measurements as best you can and determine the arbiter of success.

### 4.Consider whether the successful deployment of model results is part of data mining success. Start planning now for deployment.

### Because this project is so complicated, we have to make a good plan and arrangement for banking business.

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Time | Resources | Risk |
| Business understanding | 0.5 weeks | Analysis specialist and Advance manager of users | The bias of demand of banking business |
| Data understanding | 2 weeks | All analysts and all users | The misconceive of data |
| Data preparation | 4 weeks | Data managers of users and analysts | Missing key data |
| Data transformation | 1 week | Data transformation specialists | Losing key data |
| Data-mining method(s) selection | 3.5 weeks | Analysis specialist and Advance manager of users | Missing more efficiency Method |
| Data-mining algorithm(s) selection | 2 weeks | Analysis specialists and Advance manager of users | Choosing wrong algorithms |
| Data Mining | 2 weeks | Analysis specialists and All users | Meeting unpredictable things |
| Interpretation | 1 week | All analysts and key users | Misunderstanding conception |

## 2. Data understanding.

This dataset comes from the Bank Marketing Data Set (https://archive.ics.uci.edu/ml/datasets/bank+marketing) in the UCI Machine Learning Library [Moro et al., 2014]. These data relate to direct marketing activities of Portuguese banking institutions. These direct marketing campaigns are based on the phone. In general, the banking agency's customer service staff needs to contact the customer at least once to find out if the customer will subscribe to the bank's products (time deposits). Therefore, the task corresponding to the data set is a classification task, and the classification target is to predict whether the customer is (yes) or not (no) subscribes to the time deposit (variable y). We chose one Csv file named bank-additional-full.csv to mining : it contains all the samples (41188) and all feature inputs (20), sorted by time (from May 2008 to September 2010). Actually, we can not do more in collecting data phrase, we only able to use it in the most possible.

From the website we know that there were more than 150 properties in the beginning. We actually got the data which was performed manual feature selection from the telemarketing data obtained on the machine learning website. Some bank managers (domain expert) to define a set of related attributes, choosing from the original 150 features. Each record includes an output target, a contact result ("failed", "success"), and a candidate input variable. Input variables include telemarketing attributes (such as age, job type, contact type, previous marketing results, etc.). These records (such as the characteristics of social and economic impacts, the rate of change in unemployment rates) are collected through the Portuguese Statistics website data sources for a large number of potentially useful external data.

There are 20 input variables: age,job ,marital ,education ,default(has credit in default）, housing(has housing loan), loan(has personal loan),contact,month, day\_of\_week, duration(last contact duration, in seconds ), campaign, pdays,previous, poutcome, emp.var.rate, cons.price.idx, cons.conf.idx,euribor3m,nr.employed

Classified, the results are shown in Table 3-1:

|  |  |
| --- | --- |
| bank client data | age,job ,marital ,education ,default, housing:, loan |
| elated with the last contact of the current campaign | loan,contact,month, day\_of\_week, duration |
| other attributes | campaign, pdays,previous, poutcome |
| social and economic context attributes | emp.var.rate, cons.price.idx, cons.conf.idx,euribor3m,nr.employed |

And there is only one output variable (desired target): y - has the client subscribed a term deposit? (binary: 'yes','no')

As we can see from Figure ,compared with young people, older people are more likely to order deposits. Blue-collar workers are the least likely to subscript deposits. Retirees are the most likely to have deposits. Singles or divorcees are more likely to script deposits than married people. The education degree is also an important factor. People who have more education is more likely to order deposits. Those with credit default records basically do not order deposits. When there are houses and loans, it seems that there is no impact on whether to order deposits.



In Figure ,basic on the data from last contacting, calling by cellular is more efficiency than by telephone,and contacting in second half year is more productive than in first half year.Constant, phone in which day in a week,it seems it does not matter,every day is almost same.But the duration in every call is very important, if a client have more time on the conversation,they are more easily to order deposits,however, we could not know how long the client will be on line before we predict the success rate.So we could not use this field to predict, but we can speak more with every customer as we know it is very useful to enhance client order our product.

## 

From Figure ,we may know that the possibility of purchasing in customers who were contacted by the same salesperson is far greater than those were called by different salespersons. As for the day gap of two times contacting a customer, there is no impact on whether to order the product. It is worth noting that there are abnormal data in the interval days, is 999, according to our further analysis, 999 means that we have never contacted this customer before.

The likelihood of a customer ordering a deposit increases as the number of contacts with the customer increases during the sales activity, and customers who have previously subscribed to a deposit are more likely to re-order than other customers.

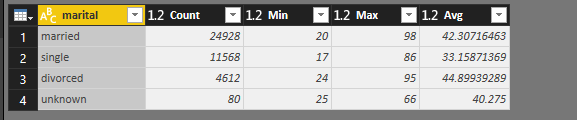


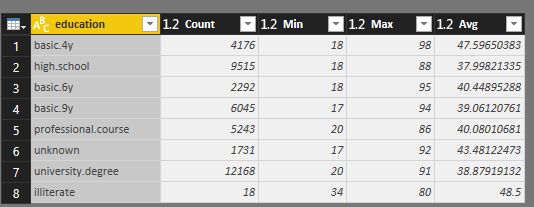
From Figure ,there are five fields to indicate social and economic context attributes:employment variation rate idx, consumer price index,consumer confidence index,euribor 3 month rate,number of employees. They are both very important for predicting weath a customer will subscript a deposit.

As we browsed these data more and more, we found some data which may be was wrong,for example,in Figure, Somebody aged 23 is retired，We think it is impossible,so we should check it again and there is a kind job named ‘admin.’ that we think it is input error, it should be ‘admin’.

## 

Also we may know before,there are some data which labeled unknown in different fields,from Figure...,these data which were labeled unknown is less than 5 percentage in our whole data set, in the next step,we can simply to discard them.

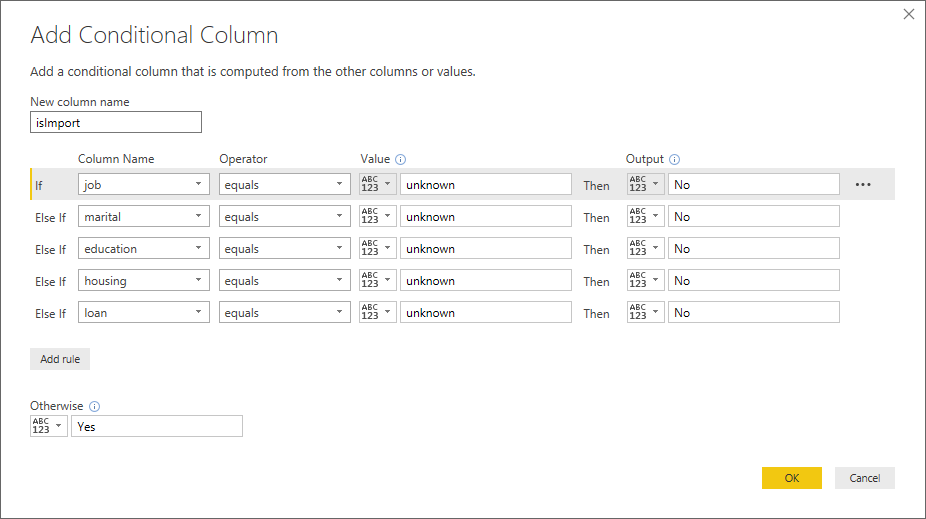


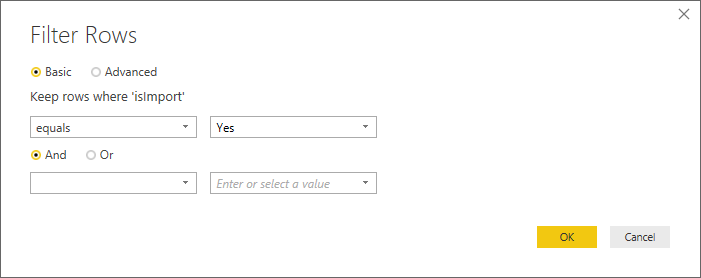


## 3.Data preparation

“Third is data cleaning and pre-processing. Basic operations include removing noise if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, and accounting for time-sequence information and known changes” (Fayyad et al., 1996)

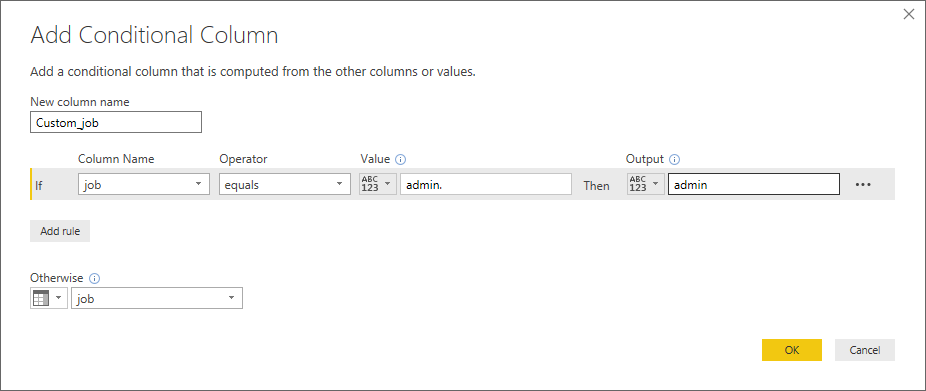
As we know there are some data which were labeled ‘unknown’, we want to discard them,so at first, we create a new column named ‘isImport’ by condition ‘if job=’unknown’ or marital=’unknown’ or education=’unknown’ then ‘No’ else ‘Yes’ in Figure . Then we filter rows by keeping rows where isImport=’Yes’ in Figure .





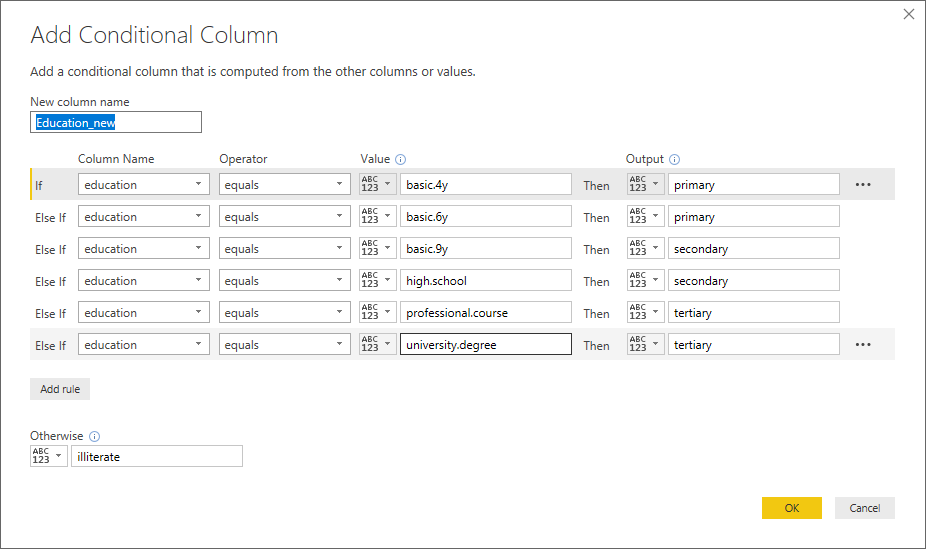
(Figure 2)

As we found there are some error data in field job in step 2, we change ‘admin.’ to ‘admin’ by three steps including add a field “custom\_job” and remove original field “job”,then rename the new field “customer\_job” to “job”,these all show in Figure .





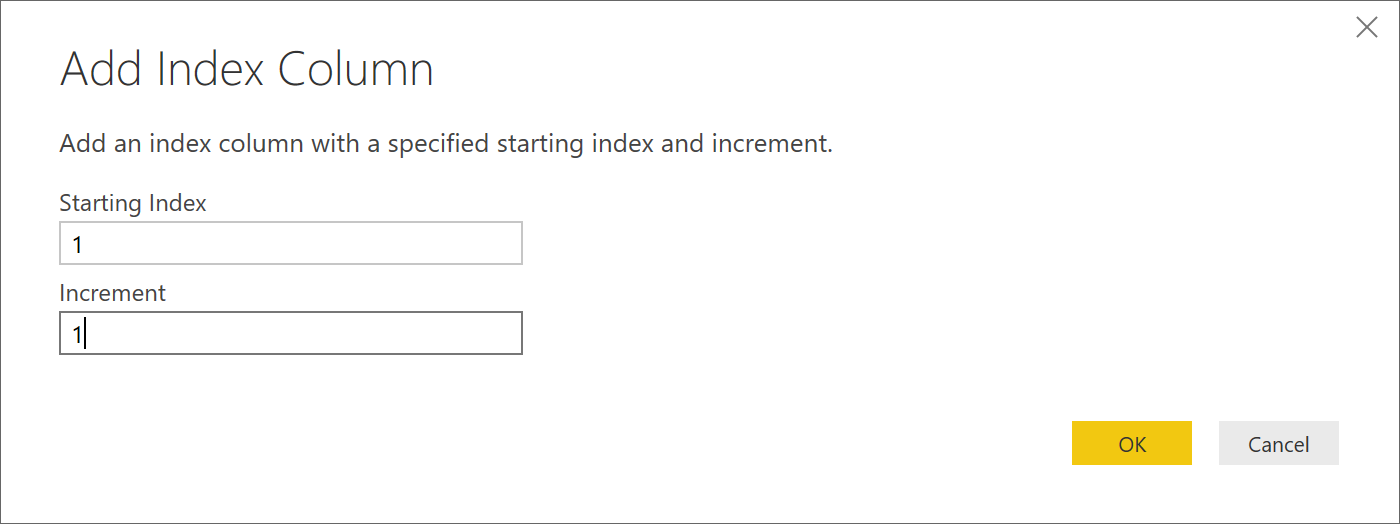
For education,there are too complicated , it is best if we merge them to three different values (‘illiterate’,‘primary’,’secondary’,’tertiary’) like most of people classify it.



From step 2,we know there are some error data related fields job and age,we should filter them;

Due to there is not a primary key in original data, we decide to add a field named RecordID by function @Index as a primary key as showing in Figure 4..

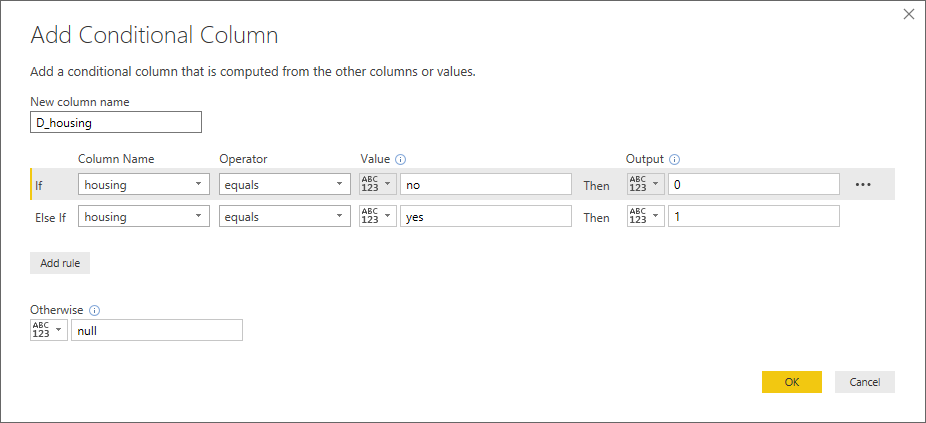
It will be convenient to do merging operation.



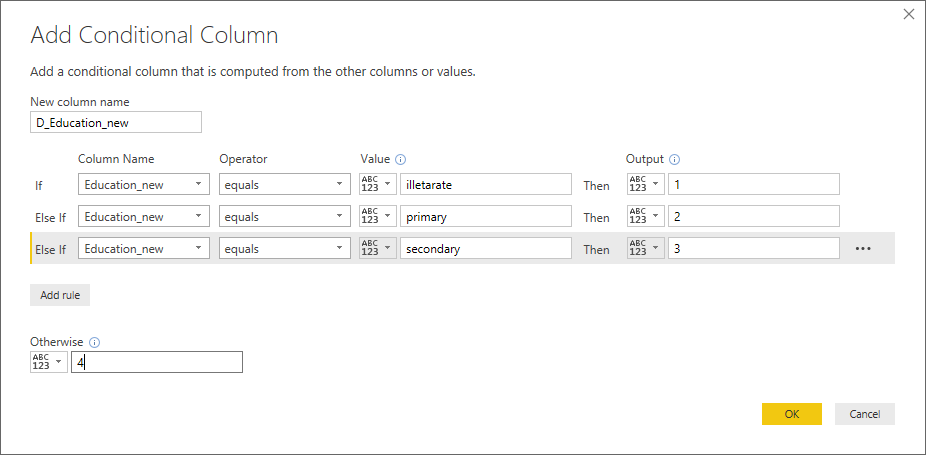
(Figure 4)

Some models only can be calculated with quantify variables,so we have to convert categorical variables to quantify variables. Actually, the categorical variables can be further divided into three kinds of quantify variables including two categorical variables,ordered categorical variables and unorder categorical variables.

In here,housing and loan are both two categorical variables,so we simpley convert them to (0=’No’,1=’Yes’,null=’unknown’)



The variable education\_new can be considered as an ordered categorical variable,we want to convert ‘illiterate’,’primary’,’secondary’,’tertiary’ to 1,2,3,4

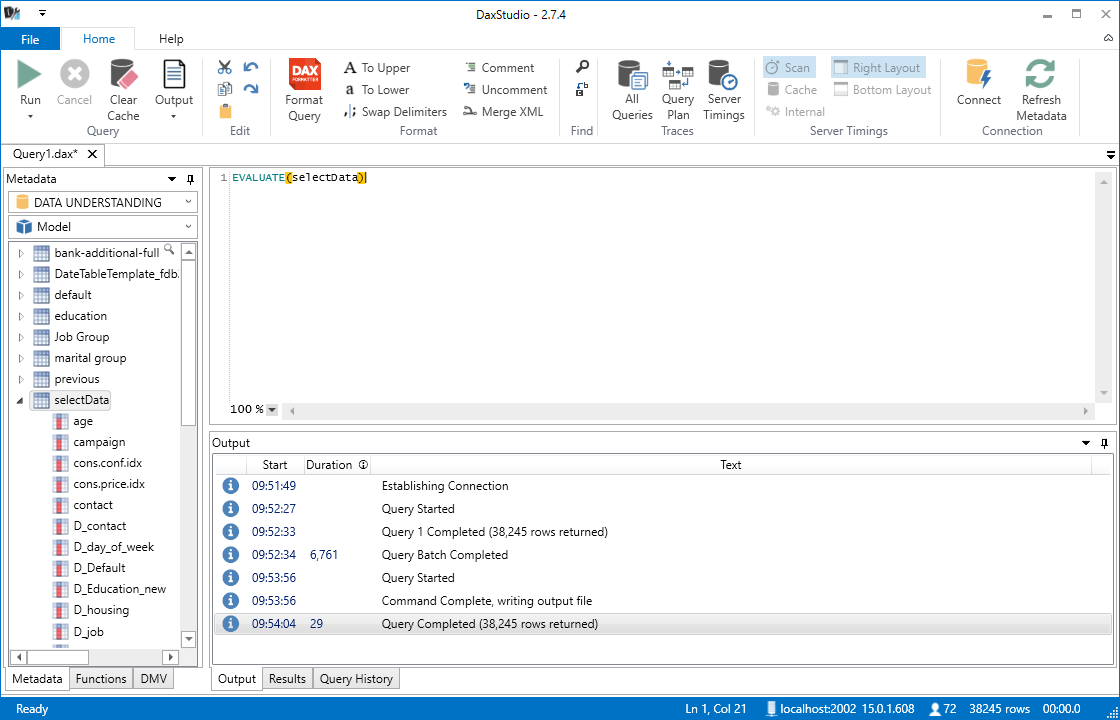


**In general, the variable marital can be thought of as an unordered categorical variable. It should be noted that encoding can be performed using dummy variables. N categories need to set n-1 dummy variables. The variable marital is divided into divorced, married, single, encoded using two dummy variables M1 and M2.**

|  |  |  |
| --- | --- | --- |
| marital | M1 | M2 |
| divorced | 0 | 0 |
| married | 1 | 0 |
| single | 0 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| FieldName | type | value | Value Remark |
| D\_default |  | 0,1,2 | 1:yes,0:no,2:unknown |
| D\_housing |  | 1,0 | 1:yes,0:no |
| D\_loan |  | 1,0 | 1:yes,0:no |
| D\_Education\_new |  | 1,2,3,4 | 1:illiterate,2:primary  ,3:secondary,4:tertiary |
| D\_job |  | 1,2,3,4,5,6,7,8,9,10 | 1:admin,2:blue-collar, 3:entrepreneur ,4:housemaid, 5:management,6:retired, 7:self-employed,8:services, 9:student,10:technician |
| D\_contact |  | 1,0 | 1:cellular,0:telephone |
| D\_month |  | 1,2,3,4,5,6,7,8,9,10,11,12 | 1:January,2:February,3:March,4:April,  5:May,6:June,7:July,8:August,  9:September,10:October,11:November,12:December |
| D\_day\_of\_week | 1 | 1,2,3,4,5 | 1:mon,2:tue,3:wed,4:thu,5:fri |
| D\_poutcome | 0 | 1,2,3 | 1:failure,2:nonexistent,3:success |

After data preparing,we use DaxStudio output “selectData” to “bank after data preparing.csv”.



## 4.Data transformation

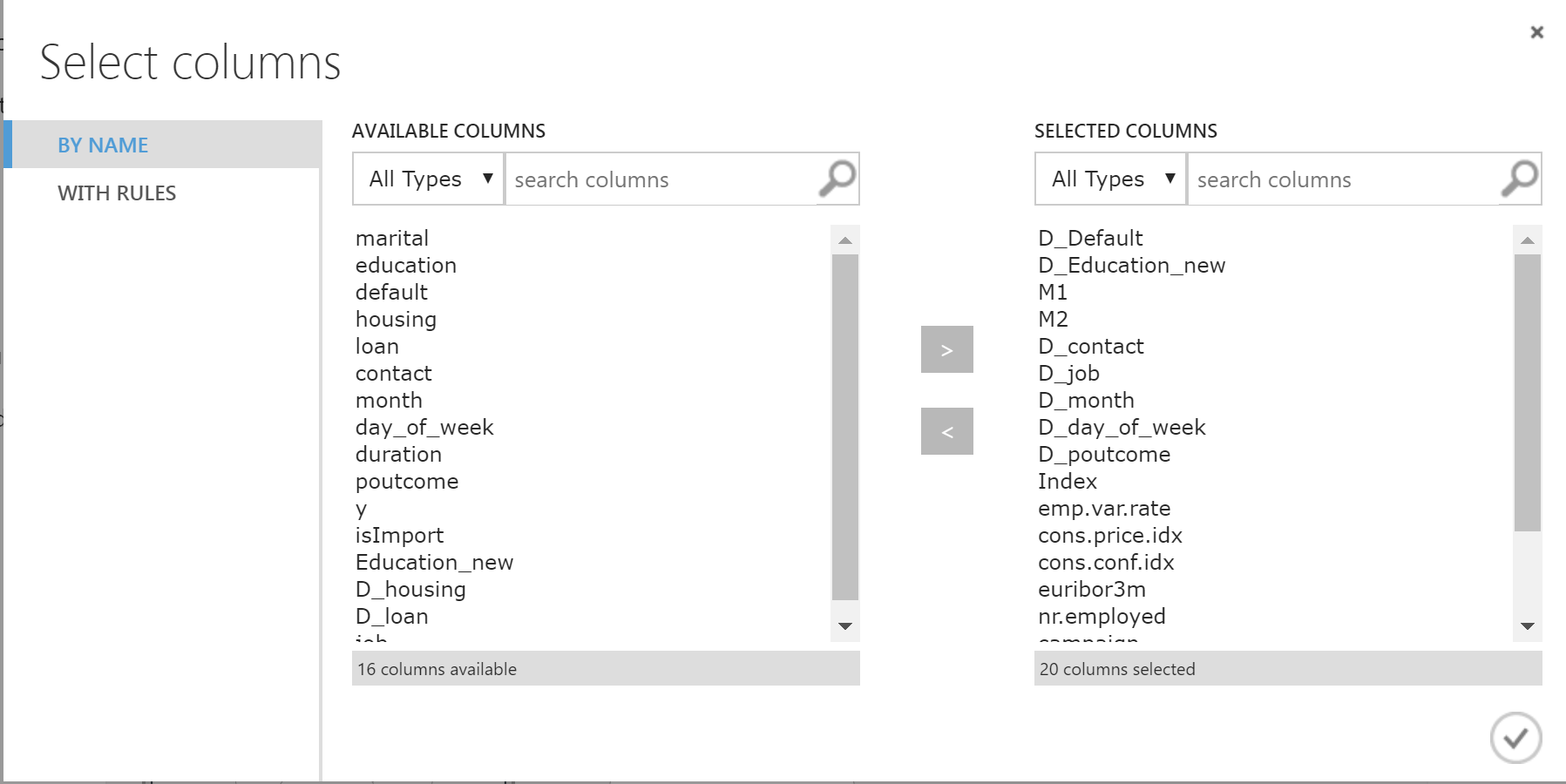
“Fourth is data reduction and projection: finding useful features to represent the data depending on the goal of the task. With dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found.” (Fayyad et al., 1996)

### 

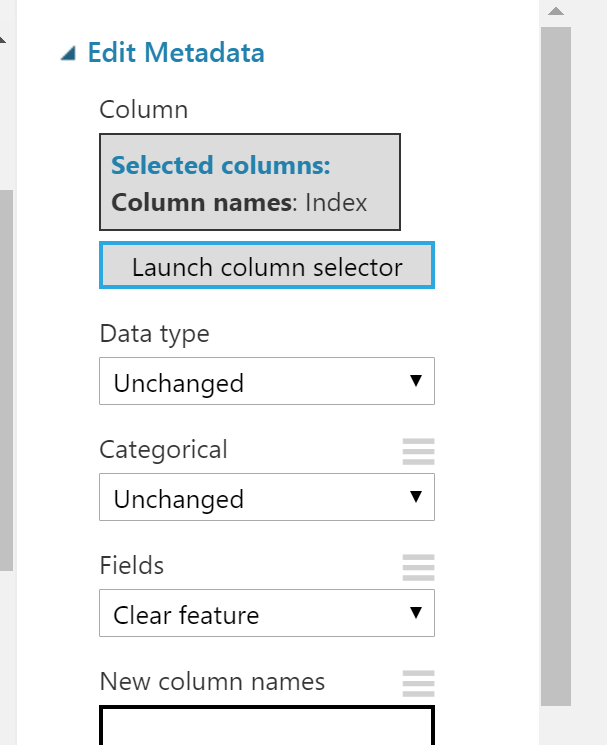
As we did in data preparing stage, we converted all of categorial fields to numeric fields, so we just need to select numeric fields to analysis and training. In additional, basing on some research in data understanding stage, we known that houses and loans did not impact on whether to order deposits.So we also can discard these two field.

For the duration, this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should be discarded as we want to have a realistic predictive model.

In conclusion, we just need to select 20 fields which were shown in Figure

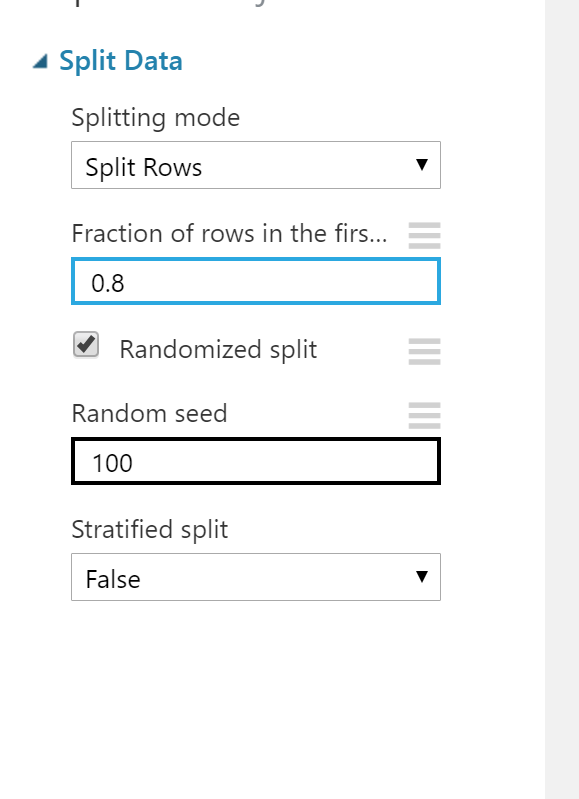


Then we add a field named “Index” to join data after prediction, so it is not a feature for prediction. We use “Edit metadata” to clear its feature.

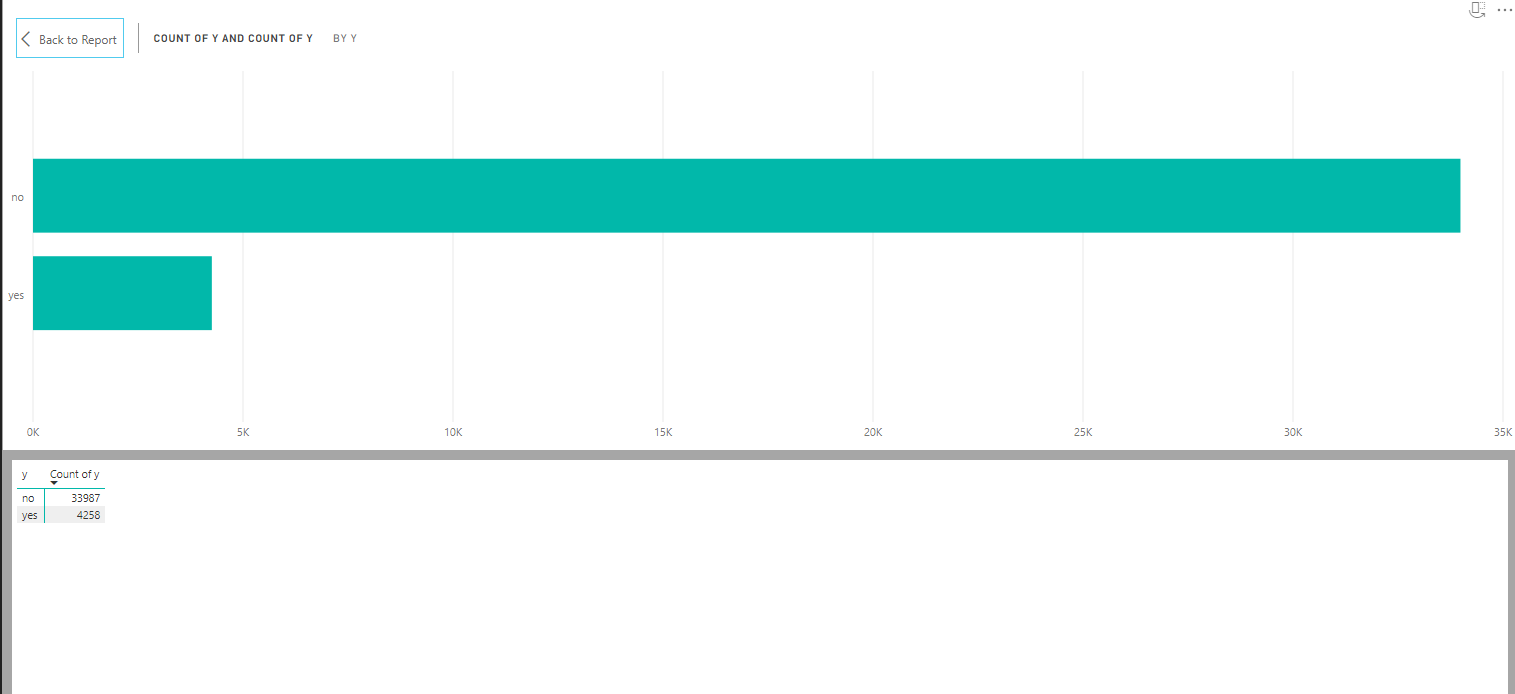


Then we should split the whole data set to two parts, one is for training and another is for testing.

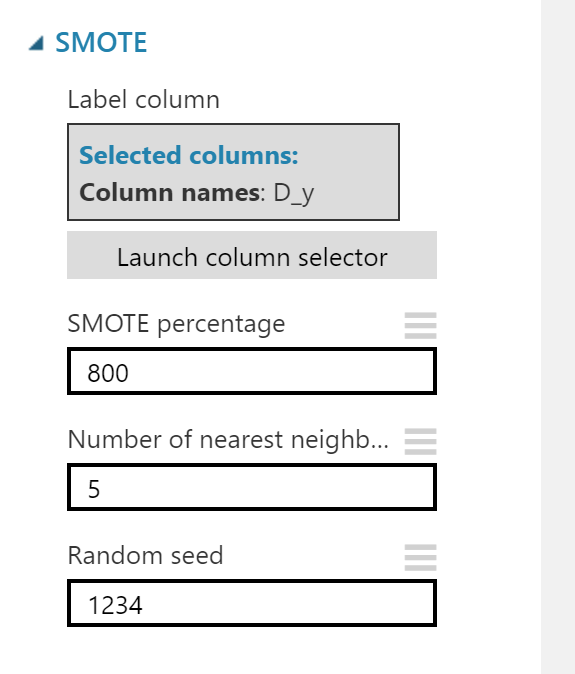
What is our decision is split 80 percentage of data for training ,and the remaining 20% of data for testing and validation.



As we can see in Figure 7 that the data is skewed (11% for Yes and 89% for No) in field Y. Skewed data may have an impact on the prediction of marketing results. It needs to be processed to be balanced. If we did not do this, the result will focus in “no” due to there are more data which ware labeled “no”, therefore the result would be bias and inaccuracy. In this kind of situation, there are two methods to deal with it, named oversampling and undersampling. Due to undersampling may lose some data features, we decided to use SMOTE to oversample our training data.

As we can see in Figure, we set SMOTE percentage to 800%, then the count of “1” and “0” in “d\_y” are almost the same in the training dataset.

see



## 5.Data-mining method(s) selection

“Fifth is matching the goals of the KDD process (step 1) to a particular data-mining method. For example, summarization, classification, regression, clustering, and so on, are described later as well as in Fayyad, Piatetsky-Shapiro, and Smyth (1996).” (Fayyad et al., 1996)

The main task we discussed before is to predict the success of a call, this is a standard classification task. So we should choose some classification models to deal with it. We focus on eight classification models which are “Multiclass Decision Forest”, ”MultiClass Decision Jungle”, ”Multiclass Logistic Regression”, ”Two-Class Bayes Point Machine”, ”Two-Class Averaged Perceptron”, ”Two-Class Boosted Decision Tree”, ”Two-Class Support Vector Machine” and ”Two-Class Locally-Deep Support Vector Machine”.

Multiclass Decision Forest works by building multiple decision trees and then voting on the most popular output classes. Voting is a form of aggregation in which each tree in the classification decision tree outputs a non-standardized frequency histogram of labels. Aggregation processes sum these histograms and normalize the results to obtain the "probability" of each tag. Trees with high predictive confidence have greater weights in the final decision of sets. Multiclass Decision Forest can capture nonlinear decision boundaries. Because Multiclass Decision Forest is effective at both computing and memory usage, it can train and predict large amounts of data. Also it can avoid noise data.

Multiclass Decision Jungle is a nonparametric model that can represent nonlinear decision boundaries. It performs the selection and classification of integration functions and are flexible in the presence of noise functions.

Multiclass Logistic Regression is a famous method in statistics, which is used to predict the probability of results, and is especially suitable for classification tasks. The algorithm predicts the probability of occurrence by fitting data to logical functions.

Two-class Bayesian Point Machine uses a Bayesian method for linear classification, which is called "Bayesian Point Machine". The algorithm effectively approximates the theoretically optimal Bayesian mean of the linear classifier (in terms of generalization performance) by selecting an "average" Bayesian point. Because Bias point machine is a Bias classification model, it is not easy to overmatch training data.

Two-Class Averaged Perceptron is an early and very simple version of the neural network. In this method, the input is classified as several possible outputs based on a linear function, and then combined with a set of weights derived from the eigenvectors - hence called "perceptrons". It is suitable for learning linear separable patterns, and neural networks (especially deep neural networks) can simulate more complex class boundaries. However, perceptrons are faster, and because they continuously process cases, perceptrons can be used for continuous training.

Two-Class Boosted Decision Tree is an ensemble learning method in which the second tree corrects the errors of the first tree, the third tree corrects the errors of the first and second trees, and so on. Prediction is based on ensemble of whole trees for prediction. Improving the decision tree is the simplest way to get the best performance in various machine learning tasks. However, they are also one of the memory-intensive learners, and current implementations store everything in memory. Therefore, the elevated decision tree model may not be able to handle very large data sets that some linear learners can handle.

Two-Class Support Vector Machine is an in-depth supervised learning method. This particular implementation is suitable for predicting two possible outcomes based on continuous or categorical variables. The SVM model is a supervised learning model that requires labeled data. During the training process, the algorithm analyzes the input data and identifies patterns in a multidimensional feature space called hyperplane. All input examples represent points in this space and map to output classes in such a way that classes are divided by the widest possible gap.

Two-Class Locally-Deep Support Vector Machine is an extremely popular and well-researched class of supervised learning models, which can be used in linear and non-linear classification tasks. In this implementation from Microsoft Azure Machine Learning, the kernel function that is used for mapping data points to feature space is specifically designed to reduce the time needed for training while maintaining most of the classification accuracy.

For the whole result, scores list in the Table 1- Eight modelling result comparing above:

|  |  |  |
| --- | --- | --- |
| Order | Modeller Name | Accuracy |
| 1 | Two-Class Boosted Decision Tree | 83.0% |
| 2 | Two-Class Locally-Deep Support Vector Machine | 81.4% |
| 2 | Multiclass Decision Forest | 81.4% |
| 4 | Multiclass Logistic Regression | 78.6% |
| 5 | Multiclass Decision Jungle | 71.1% |
| 6 | Two-Class Bayes Point Machine | 71.1% |
| 7 | Two-Class Averaged Perceptron | 68.6% |
| 8 | Two-Class Support Vector Machine | 43.2% |

(Table 1-Eight modelling result comparing)

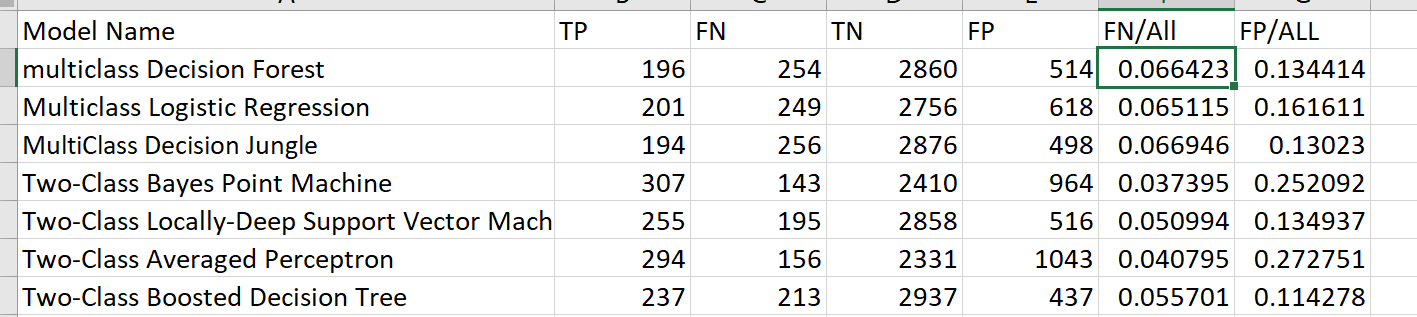
Basing on Table 1- Eight modelling result comparing, obviously “Two-Class Support Vector Mach” got a very low result, it looks it totally did not match this task, so we decided to choose other seven methods do the next step research.

## 6.Data-mining algorithm(s) selection

“Sixth is exploratory analysis and model and hypothesis selection: choosing the datamining algorithm(s) and selecting method(s) to be used for searching for data patterns. This process includes deciding which models and parameters might be appropriate (for example, models of categorical data are different than models of vectors over the reals) and matching a particular data-mining method with the overall criteria of the KDD process (for example, the end user might be more interested in understanding the model than its predictive capabilities).” (Fayyad et al., 1996)

For this task, as we know, there are only 10% of customers will be attracted

by marketing activities. So comparing to the correct rate of the whole data, the correct rate of customers who labelled success is more important. So we decide to look the different between these parts each modeling got. The total of results from different models list on Figure



So basically if we want to focus more on minimising False Negatives, we would want our Recall to be as close to 100% as possible without precision being too bad and if we want to focus on minimising False positives, then our focus should be to make Precision as close to 100% as possible

It is clear that recall gives us information about a classifier’s performance with respect to false negatives (how many did we miss), while precisiongives us information about its performance with respect to false positives(how many did we caught).

From Table-2 corrent on y=”yes” Comparing ,“Two-Class Bayes Point Machine”,” Two-Class Averaged Perceptron” and “Two-Class Locally-Deep Support Vector Machine“ got the best three score on “YES”, they got 68.2%,65.3% and 56.7% respectively. So we decided these three methods will be able to enter the next step.

|  |  |  |
| --- | --- | --- |
| Order | Modeller Name | Recall(The Rate of correct on y=”Yes”) |
| 1 | Two-Class Bayes Point Machine | 68.2% |
| 2 | Two-Class Averaged Perceptron | 65.3% |
| 3 | Two-Class Locally-Deep Support Vector Machine | 56.7% |
| 4 | Two-Class Boosted Decision Tree | 52.7% |
| 5 | Multiclass Logistic Regression | 44.7% |
| 6 | Multiclass Decision Forest | **43.6%** |
| 7 | Multiclass Decision Jungle | 43.2% |

(Table-2 corrent on y=”yes” Comparing)

Specificity is a measure that tells us what proportion of patients that did NOT have cancer, were predicted by the model as non-cancerous. The actual negatives (People actually NOT having cancer are FP and TN) and the people diagnosed by us not having cancer are TN. (Note: FP is included because the Person did NOT actually have cancer even though the model predicted otherwise).

Specificity is the exact opposite of Recall.

From Table-3 Incorrent on y=”no” Comparing, “Two-Class Boosted Decision Tree“, “Multiclass Decision Jungle” and “Multiclass Decision Forest” got the best three score on original “NO”, they only miss 13,0%,14.8% and 15.2% respectively. It means If we add these prediction data to the waiting phoning pool, it will not increase largely the number of failures. So we decided these three methods also will be able to enter the next step.

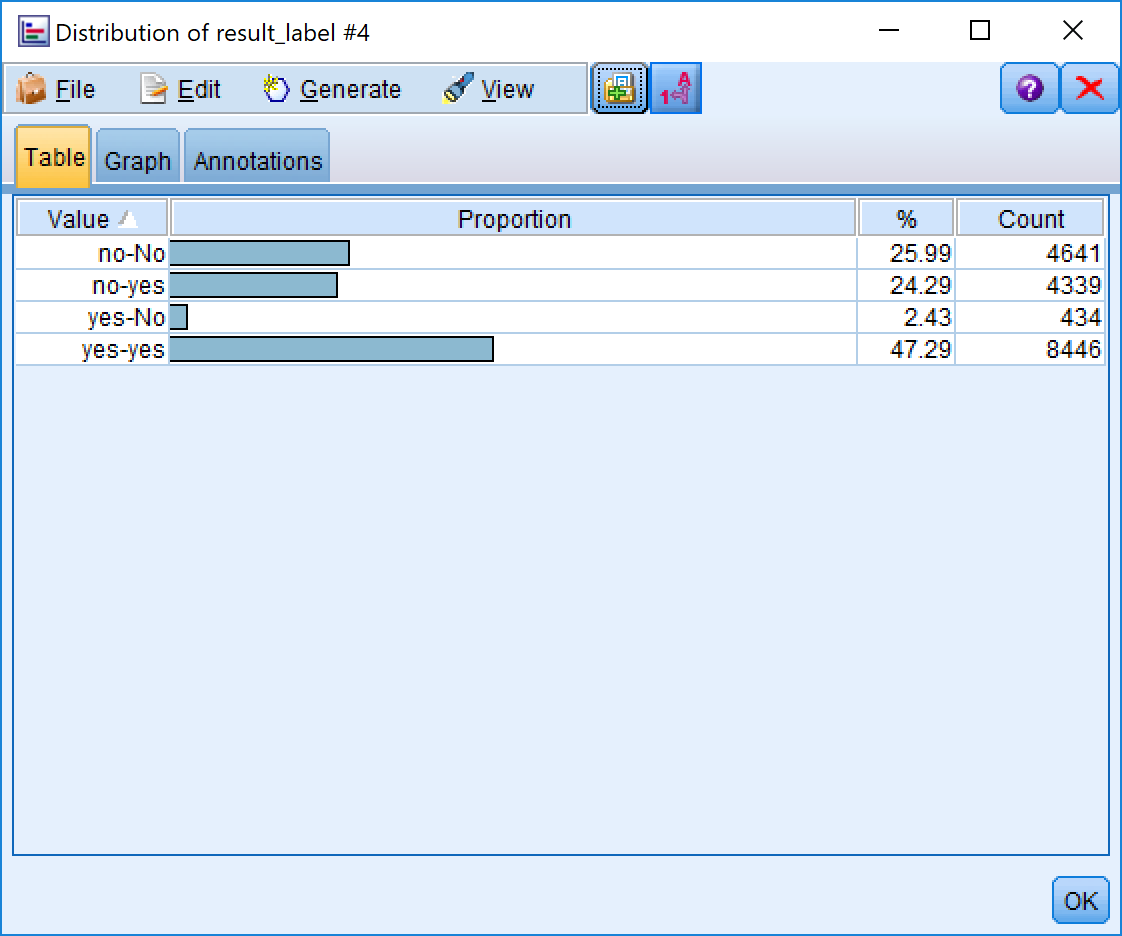
|  |  |  |
| --- | --- | --- |
| Order | Modeller Name | Specificity (The Rate of correct On y=”no”) |
| 1 | Two-Class Boosted Decision Tree | 87.0% |
| 2 | Multiclass Decision Jungle | 85.2% |
| 3 | Multiclass Decision Forest | 84.8% |
| 4 | Two-Class Locally-Deep Support Vector Machine | 84.7% |
| 5 | Multiclass Logistic Regression | 81.7% |
| 6 | Two-Class Bayes Point Machine | 71.4% |
| 7 | Two-Class Averaged Perceptron | 69.0% |

(Table-3 Incorrent on y=”no” Comparing)

So Totally, six Models are still in the pool we want to choose. They are ,“Two-Class Bayes Point Machine”,” Two-Class Averaged Perceptron” , “Two-Class Locally-Deep Support Vector Machine”,“Two-Class Boosted Decision Tree“, “Multiclass Decision Jungle” and “Multiclass Decision Forest”. But there is also no any modelling which match the original target in the beginning of this report. Due to the accuracy on Yes is in 60%-70%,lower than our target, and the accuracy on No is also between 80-90%,higher than our target. So we could do something to improve the accuracy on Yes by decreasing the accuracy on No.

It is our final decision to combine the six models result, if any of these six models predict it is Yes, we predict it is “Yes”, in other words, if all of these models predict a customer will answer NO, we set it is No, otherwise it will be Yes.

Let us have a look at the final data in Figure 16.



(Figure 16)

## 7.Data Mining

“Seventh is data mining: searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression, and clustering. The user can significantly aid the data-mining method by correctly performing the preceding steps.” (Fayyad et al., 1996) This is, of course, the flashy part of data mining, where sophisticated analysis methods are used to extract information from the data.

As we all know, in the field of machine learning, it is very important to divide the training set verification set and test set reasonably. How to divide a large data set into training set and test set is a key factor in machine learning. The test set must satisfy the following two conditions:

The scale is large enough to produce statistically significant results.

It can represent the entire data set. In other words, the characteristics of the selected test set should be the same as that of the training set.

Assuming that the test suite satisfies these two conditions, the goal is to create a model that can be well generalized to new data. Our test set acts as a proxy for new data.

In the past, when people used machine learning methods, the training set and test set were usually classified as 7:3..

If there is a validation set, it is delimited as 6:2:2. This delimitation is really scientific, when the amount of data is small (10,000 levels and below).

We currently have about 40,000 records, so we decided to split the data by 9:1, 90% for training, 10% for testing, that is, about 4,000 rows for test certification.

## Because the accuracy of each model is lower than we expected, so we decided to use Tune Model Hyperparameters

to get the best model parameters, and use more data to verify the model parameters are appropriate, so in the training set data, we split according to the 8:2 ratio, the appropriate increase in cross-validation set data ratio.

As we can see from Figure 17, we get a very good result from putting the whole datasets into this data mining models. Data from 45% of the total is marked as no-no, indicating that the original data is No, and all five models are calculated as No. For this part of the customers, we do not need to make a phone call and then from the total of 0.59% of the data. Being marked as “YES-No” indicates that if we call out according to this forecast data, we will lose 0.59% of potential orders. In conclusion, we reduced our out-of-pocket workload by 46.46%, lost 0.59% accuracy, and increased the effective rate of exhalation from 12% to 26.32%, which is consistent with and greatly exceeds our default target, so there is no doubt this project was successful.

In order to simplify the analysis, figure 4-6 gives decision rules with less decision level. The rules are as follows:

(1) if the number of employees is less than 5087, and the type of communication is mobile phone or telephone, then the customer is not.

Thirty-six

Master degree thesis of Huazhong University of Science and Technology

Subscribe to time deposits;

(2) if the number of employees is less than 5087 and the type of communication is unknown, the number of contacts in activities is more than 3.

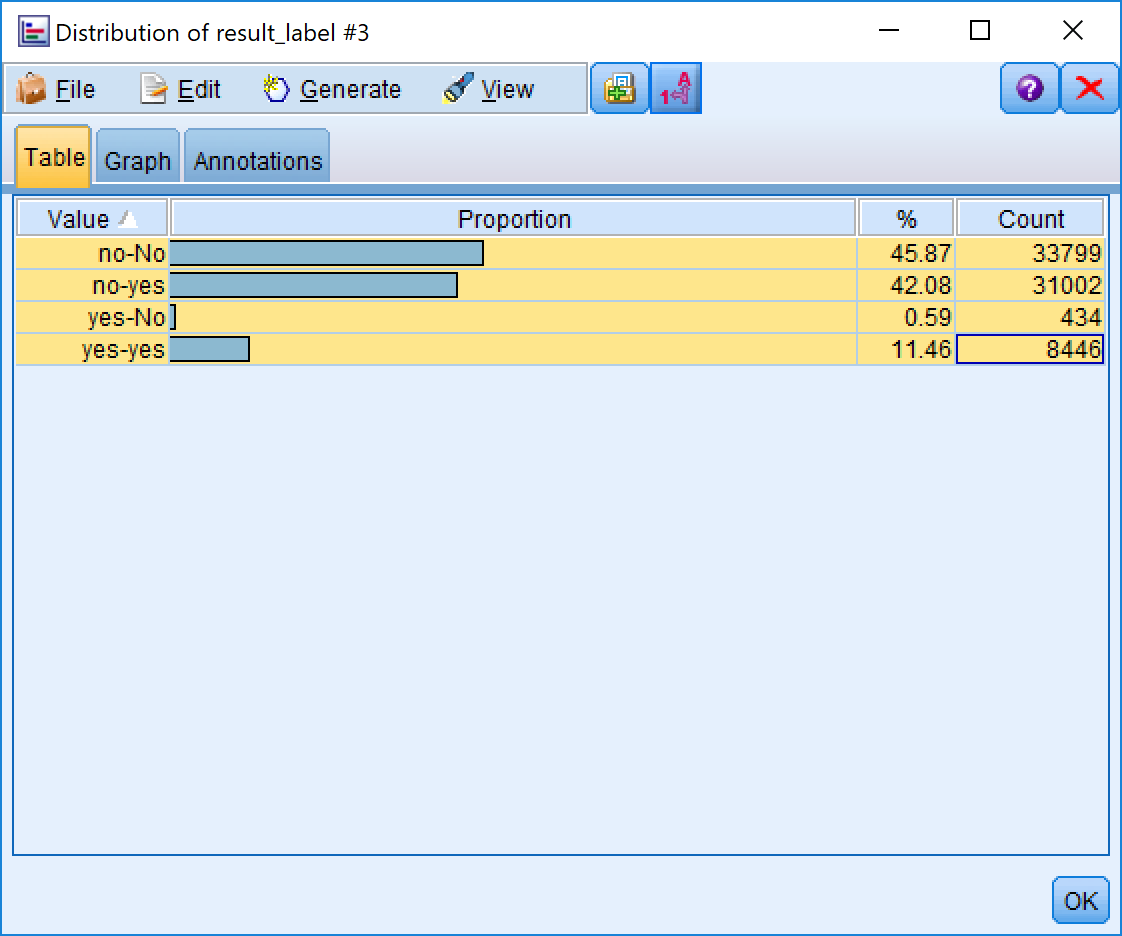
People, then customers will subscribe to time deposits.

(3) if the number of employees is greater than 5087, the duration will be between 411 and 632, and the number of people involved in the activity will be more than 632.

More than 3, then customers will not subscribe to time deposits.

(4) if the number of employees is greater than 5087 and the duration is greater than 632, the number of contacts in activities is less than 3.

Customers will subscribe to time deposits.



(Figure 17)

## 8.Interpretation

“Eighth is interpreting mined patterns, possibly returning to any of steps 1 through 7 for further iteration. This step can also involve visualization of the extracted patterns and models or visualization of the data given the extracted models.” (Fayyad et al., 1996) We assess and evaluate the models and the results and their reliability. “You are ready to evaluate how the data mining results can help you to achieve your objectives.” (SPSS, 2007)

At present, banks do not filter customers in marketing, directly calling all customers in the database. However, in the future there can be changes in the choice strategy of bank customers. For example, imagine a scenario where a telephone sales manager is asked to halve the number of customer contacts. Without the idea of a data-driven model, phone sales will reach only 50% of the expected potential users, where the use of neural network models will allow for about 97% of the response, benefiting from an increase of 47 percentage points for successful subscriptions. This result proves that the data mining model, which allows managers to improve efficiency by reducing costs, still achieves a large part of the success. By choosing the most likely customers, bank telemarketing managers can create more value by improving activity efficiency, reducing customer contact costs [45]. The classification model is a valuable tool to support telemarketing managers in selecting customers.

## Despite the result we got first, we decide to use Tune Model Hyperparameters

, we think it will give us a good suggestion on parameters settings. It took more time to run the whole model, fortunately, the result improved a lot, as shown in Figure.

In the future, timely update the telephone marketing database, the best model in this paper will be applied to the ground, select customers, and carry out personalized marketing, to create more value for banks. 2. Making full use of highly relevant attributes, two knowledge extraction techniques are also applied to model, sensitivity analysis and rule extraction. Sensitivity analysis is used to input the most relevant features in neural network. The results of decision rules and sensitivity analysis show that the top ten attributes are consistent. In banking, these attributes provide valuable information for telemarketing managers.

The number of employees is considered the most relevant attribute, relative to 21% of the importance. The number of employees and the eighth most relevant quarterly indicator attribute (rate of employment change), these social indicators also have an important impact on marketing results.

Next comes the type of communication, duration, and number of contacts in the activity, which are linked to past marketing activities. It shows that spending a lot of time in past activities increases the likelihood of success.

Although customer attributes (including age, marital status, etc.) are specific individuals, they are considered to be rarely relevant, this does not mean that these types of attributes generally have little impact on marketing success.

When considering the eighth attribute of the Euro Interbank Offer Rate, one might think that a lower Euro Interbank Offer Rate would lead to a lower savings rate. After 2008, however, inter-bank lending rate savings in the eurozone have been declining and customer savings rates have risen. This apparent contradiction may be due to customers' perception of real economic recession and social depression. Consumers may feel the need to consider increasing savings for the future rather than spending money immediately on the products or services they need. Increasing this attribute means increasing the probability of subscription deposits.

## References

1. S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

2.S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS.

3. Langley, A., Mintzberg, H., Pitcher, P., Posada, E., & Saint-Macary, J.Opening up decision making: The view from the black stool. organization Science, 6(3), pp. 260-279, 1995.