ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Econometrics and Operations Research

Bachelor Thesis Business Analytics and Quantitative Marketing

SPEEDY SUCCESS

A data-driven approach to select the best set of potential telemarketing customers for selling long-term bank deposits

Author: P.E. (Pepijn) van der Leije (455987)

Supervisor: dr. M. (Mikhail) Zhelonkin

Second assessor: U. (Utku) Karaca

Date final version: 5 July 2020

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

ABSTRACT

In this study, we predict the success of telemarketing calls for selling long-term bank deposits using five classification methods: the decision tree, the logistic regression, the support vector machine, the artificial neural network and the random forest. In addition, we predict the contact time of telemarketing calls using three regression techniques: the logarithmic transformed ordinary least squares regression, the poisson regression and the negative-binomial regression. Applying data of 41,176 phone contacts collected from a Portuguese retail bank from the period May 2008 to November 2010, we discover that the artificial neural network and the negative-binomial regression are the two best performing methods to predict the success and contact time. Moreover, we detect which characteristics influence the success rate and contact time. The employment rate, the number of days since the last call for any other campaign and the Euribor rate are most positive influential for the success rate. Furthermore, potential customers with loans in delay and an increase in the Euribor rate have the largest significant negative effects on the contact time. Eventually, we create a self-made function to form a list of 868 potential customers where the first potential customer on the list has the highest success probability per second and thus, is most attractive to call for the bank.

Contents

1 Introduction **1**

2 Related Literature **3**

3 Data **4**

4 Methodology **7**

4.1 Success Prediction 7

4.1.1 Decision Tree 8

4.1.2 Logistic Regression 8

4.1.3 Support Vector Machine 9

4.1.4 Artificial Neural Network 9

4.1.5 Random Forest 10

4.2 Contact Time Prediction 10

4.2.1 Logarithmic Transformed Ordinary Least Squares Regression 11

4.2.2 Poisson Regression 11

4.2.3 Negative-Binomial Regression 12

4.3 Characteristics and the best set of potential costumers 13

4.3.1 Success and contact time characteristics 13

4.3.2 The best set of potential costumers 13

5 Results **14**

5.1 Success results 14

5.2 Contact time results 15

5.3 Characteristics results and the best set of potential costumers 16

6 Conclusion and Discussion **19**

6.1 Conclusion 19

6.2 Discussion and Outlook 20

References **21**

A Appendix **23**

B Appendix **23**

C Appendix **24**

D Appendix **25**

1 Introduction

One way to advance business is by using marketing campaigns. Almost all companies advertise and promote their products. Nowadays, one of the most popular forms is telemarketing. Telemarketing is a direct advertising approach towards potential costumers through telephone communication. Companies intend to contact the best set of potential costumers. In other words, the ones who are most likely to buy their product. It is valuable to constrict the range of potential customers and to know their characteristics. This may increase the success rate as well as efficiently reduce the marketing costs. Besides increasing the success rate, it is interesting to know how much time and effort a company needs to put in before a product is bought. Therefore, to advise companies as efficiently as possible, we focus on the probability of a product to be bought and, in addition, the time it takes to sell it.

In this study, we predict the success of telemarketing calls for selling long-term bank deposits, which has already been researched by Moro, Cortez and Rita (2014). In addition, we forecast the time it takes to sell these long-term deposits. Finally, we determine the influence of potential customer characteristics and present the best set of potential customers. Thus, we pose the following research questions:

- *To what extent are various classification methods capable of examining the success of telemarketing calls for selling long-term bank deposits?*
- *And, to what extent are regression methods capable of forecasting the contact time of telemarketing calls for selling long-term bank deposits?*
- *And, which characteristics influence the success rate and contact time of telemarketing calls for selling long-term bank deposits?*
- *Finally, what is the best set of potential customers for selling long-term bank deposits?*

Using data of 41,176 phone contacts from a Portuguese retail bank from the period May 2008 to November 2010 provided by UCI Machine Repository, we analyse 21 input features related to personal, contact, historic and economic information. For evaluation purposes, we split the data in two subsets. The training set is executed up to July 2010 and the test set includes the most recent phone contacts from August 2010 to November 2010.

To examine the success of telemarketing calls and detect the potential costumers who are most likely to buy a long-term bank deposit, we apply five classification methods: the decision tree (DT), the logistic regression (LR), the support vector machine (SVM), the artificial neural network (ANN) and the random forest (RF). Here, we use two metrics to compare the method performances: the area under the receiver operating characteristic curve (AUC) and area under the LIFT cumulative curve (ALIFT). To predict the contact time of telemarketing calls, we apply a logarithmic transformed ordinary least squares regression (OLS) and two count data regressions: the poisson regression (POI) and the negative-binomial regression (NBM). We select the best contact time predicting method by minimizing the mean squared prediction error (MSPE).

To determine which characteristics influence the success of telemarketing calls, we apply the best performing classification method, the artificial neural network (AUC = 0.757, ALIFT = 0.632), and a data-based sensitivity analysis algorithm. Here, we notice that the employment rate, the number of days since the last call for any other campaign and the Euribor rate have the largest positive influence. Moreover, we use the sign and significance of the best resulting regression method, the negative-binomial regression (MSPE = 551,313), to discover which characteristics influence the contact time of telemarketing calls. We find that potential customers with loans in delay and an increase in the Euribor rate have the largest significant negative effects on the contact time.

We use the artificial neural network and the negative-binomial regression to predict the success rate and contact time of all 868 potential customers from the test set. Afterwards, we create a self-made function to select the best set of potential customers based on the highest success probability per second, where we try to make the situation as realistic as possible. By doing this, we are able to form a list of 868 potential customers where the first potential customer on the list has the highest success probability per second (0.1213) and thus, is most attractive to call for the bank.

The remaining part of this study is structured as follows. We discuss the theoretical impact evaluation regarding success of telemarketing calls based on the existing literature in Section 2. In Section 3, we describe the data in more detail, followed by methods and techniques used to answer the research questions in Section 4. Then, we describe and compare the obtained results in Section 5. Finally, in section 6 we present the conclusions, as well as some limitations and suggestions for further research.

2 Related Literature

The popularity of big data has been able to grow vehemently due to the developments of the modern information technology. In many different fields big data is useful for decision makings and predictions, likewise in the banking sector (Chen, Han, & Yu, 1996). For instance, Moro et al. (2014) applied different data-driven methods to improve the success of telemarketing calls for selling long-term bank deposits. With a data set of 22 input features, they analyzed the performance of four methods on an evaluation set, applying a holdout estimation and a rolling window scheme. Eventually, the artificial neural network method obtained the best outcome in both situations, permitting to attain 79 percent of successful sales while only contacting half of the potential costumers. In addition, they applied a sensitivity analysis to the artificial neural network method to disclose the most influential input features.

According to Jaing (2018), the logistic regression method explores the relationship between the success of telemarketing and the input features best. With a data set of 21 input features, they compared the logistic regression to four methods: the bayes, the support vector machine, the artificial neural network and the decision tree. Eventually, the logistic regression attained the highest accuracy of 92.03%. Subsequently, Selma (2020) continued with the artificial neural network method. With a data set of 41,188 phone calls and 21 input features, they attained an accuracy of 98.93% and were able to outperform the results of Moro et al. (2014). However, it should be stressed that, in all three studies, some input features contain hindsight information such as the number of contacts for a campaign, the direction of a call and the time of a phone contact. As this information is unknown in advance of the phone contact, it is not practicable for a marketeer to apply this in a prediction method.

Several other studies compared the results of different classification methods in other research topics. They discovered contrasting results. For example, Cortez, Cerdeira, Almeida, Matos and Reis (2009) experienced that the support vector machine performed better than the artificial neural network for modeling wine preferences. Likewise, this was also the case in Delen (2010) who examined student retention management. According to Olson, Delen and Meng (2012), the decision tree outperformed both the support vector machine and the artificial neural network for bankruptcy prediction. In summary, classification methods outperform each other on different type of fields.

Therefore, we compare the four methods (decision tree, logistic regression, support vector machine and artificial neural nnetwork) applied by Moro et al. (2014) and Jaing (2018) in our study to conclude which method fits the success rate problem best. In addition, we introduce a fifth method to predict the success of telemarketing calls, the random forest, based on the results of Moro et al. (2014). In their study, they first use a holdout estimation and afterwards a rolling window scheme to evaluate the performance. While the decision tree performs 10% worse than the artificial neural network and is the worst method in the first evaluation, it is only 3% worse than the artificial neural network and is almost the second best method in the rolling window scheme estimation. Subsequently, according to Ali, Khan, Ahmad and Maqsood (2012) the random forest provides better outcomes than the decision tree for a large data set. Therefore, we reason that tree algorithm methods might be good techniques for realistic environment evaluation and thus, it is of interest how the random forest predicts the success of telemarketing.

Even though Moro et al. (2014) and Jaing (2018) did find an approach to improve the success of telemarketing calls for selling long-term bank deposits, the contact time of a call has not been taken into account. It is in the best interest of companies to reduce the time while the success rate remains the same. For example, a potential customer with a success probability of 80% and a contact time of 2,000 second is less interesting than five potential customer who all have a probability of 75% success and a contact time of 200 seconds per phone contact. Thus, to improve the advice for the bank even further, we focus on the contact time of each call as well as the success rate probability.

3 Data

The data set consists of 41,188 phone contacts from a Portuguese retail bank from the period May 2008 to November 2010 provided by UCI Machine Repository[1]. A phone contact is a telephone conversation with a potential customer of which we possess various original input features and four additional input features created by ourselves. Table 1 shows the description of the input features. The first dependent variable is a dummy variable which indicates whether the telemarketing call results into a long-term deposit sale or not. The second dependent variable is the numerical contact time variable, which indicates the contact time of a campaign. Furthermore, each potential costumer has 21 input features which are related to personal, contact, historic and economic information. It

[1] https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

is important to mention that all these 21 characteristics are known before the contact. Besides, two extra input features contain information about the number of contacts for a campaign and the time of the call. Since, this hindsight information is unknown before the contact, it is not accessible for a marketeer to apply this in a prediction method. A more detailed analysis of the input features is shown in Table 8 and Table 9 in Appendix A.

Table 1: Input features description

		name	description
dependent variables			
		y$^{(c)}$	Specify whether the call results into a deposit sale
	$\star$	contact.time$^{(n)}$	Total contact time of a campaign
personal information			
		age$^{(n)}$	Age in years of the costumer
		job$^{(c)}$	Type of job with 12 categorical possibilities
		marital$^{(c)}$	Marital status with 4 categorical possibilities
		education$^{(c)}$	Completed education with 8 categorical possibilities
		default$^{(c)}$	Specify whether the customer has loans in delay
		housing$^{(c)}$	Specify whether the customer has a mortgage account
		loan$^{(c)}$	Specify whether the customer has a personal credit
contact information			
		contact$^{(c)}$	Communication type with 2 categorical possibilities
	$\star$	year.2008$^{(c)}$	Specify whether the call was made in 2008
	$\star$	year.2009$^{(c)}$	Specify whether the call was made in 2009
	$\star$	year.2010$^{(c)}$	Specify whether the call was made in 2010
		month$^{(c)}$	Last contact month with 12 categorical possibilities
		day$^{(c)}$	Last contact day with 5 categorical possibilities
historic information			
		pdays$^{(n)}$	Number of days since the last call for any other campaign
		previous$^{(n)}$	Total number of past calls
		poutcome$^{(c)}$	Result of the previous campaign with 3 categorical options
economic information			
		emp.var.rate$^{(n)}$	Quarterly data of the employment variation rate
		cons.price.idx$^{(n)}$	Monthly data of the consumer price index
		cons.conf.idx$^{(n)}$	Monthly data of the consumer confidence index
		euribor3m$^{(n)}$	Daily data of the Euribor 3 month rate
		nr.employed$^{(n)}$	Quarterly number of employees in Portugal in thousands
hindsight information			
		campaign$^{(n)}$	Number of contacts for the same campaign
		duration$^{(n)}$	Contact duration in seconds for most recent call

Notes. From top to bottom: The dependent variables are a dummy variable y which indicates whether the telemarketing call results into a long-term deposit sale and a numeric *contact.time* variable which indicates the contact time of a campaign. The 21 input features, which are known before a call, are related to personal, contact, historic and economic information. Furthermore, the data set consists of two input features which are classified as hindsight information, because they are unknown before the contact. Moreover, $^{(c)}$ and $^{(n)}$ indicate whether the variable is categorical or numerical, respectively. Finally, the four $\star$ variables are not from the original data set and is discussed in further detail in the data section.

There are 12 duplicated rows in the original data set. Since the probability that two potential costumers have exactly the same 21 characteristics and an identical contact time is negligible, we delete these double observations and continue with 41,176 phone contacts. For evaluation purposes, we split the data into two subsets. The training data is executed from May 2008 up to July 2010 and the test data includes the most recent phone contacts from August 2010 to November 2010. This results in a training set and a test set of respectively 40,308 and 868 phone contacts.

To improve the original data set, we include four additional variables. First, a new dependent variable is introduced to predict the contact time of each telemarketing call. As we know the duration time in seconds of most recent call and the number of contacts for the same campaign, we generate the contact time of a campaign by multiplying those variables. The distribution of the contact time is positively skewed. A plot of this, as well as a plot of the logarithm, is shown in Figure 3a and Figure 3b in Appendix B, respectively. It should be noted that this is not the realistic contact time of a campaign, because not all previous calls for the same campaign have the same duration time as the most recent call. However, in this study we assume that it is approximately equivalent.

Second, the original data set consists only of information about the month and day of a phone contact. Here, the classification methods interpret no difference between, for example, May 2008 and May 2010 because 'year' is not included. However, 2008 was in the beginning of the great recession and 2010 at the end, which might influence the number of sales. Table 2 shows the absolute frequency and percentage frequency of the results of phone contacts for the long-term deposit sales. While the percentage of success was less than 5% in 2008, the successful sales' percentage of 2010 is more than 50%. As the original data set is ordered from May 2008 to November 2010, we are capable of adding the 'year' of a phone contact into the data set by the use of three dummy variables.

Table 2: Frequency of phone contact results

	Absolute freq.	Percentage freq.
2008	27,682	
No	26,343	95.16%
Yes	1,339	4.84%
2009	11,436	
No	9,209	80.53%
Yes	2,227	19.47%
2010	2,058	
No	985	47.86%
Yes	1,073	52.14%
Total	41,176	

Notes. The absolute frequency and percentage frequency of the results of phone contacts for the long-term deposit sales are shown. Here, 'yes' indicates that the call results into a deposit sale. The total number of phone contacts is 41,176 and the contacts are not evenly distributed over the years.

4 Methodology

4.1 Success Prediction

We use classification techniques to predict the success of telemarketing calls for selling long-term bank deposits. The purpose of a classification method is to determine the class membership of y_{n+1} with characteristics x_{n+1} based on previous data set $S = (x_1^{'}, y_1), ..., (x_n^{'}, y_n)$. In this study, y_i is a binary dependent variable, where 0 indicates no sale deposit and 1 indicates a sale. The x_i are 21-dimensional vectors that contain information about the potential customers. We describe the relation between x_i and y_i by a probability distribution $P(x_i, y_i)$. Eventually, we are interested in $P(y_i \mid x_i)$, so we may predict the probability of a deposit sale of potential customers, given their input features.

We apply the following five classification methods: the decision tree, the logistic regression, the support vector machine, the artificial neural network and the random forest. A short explanation of the five methods, including their advantages and disadvantages, is provided below. The classification codes are programmed in R with the package 'rminer' (Cortez, 2020). To evaluate the performance of the different methods, we use the same two techniques as Moro et al. (2014). First, we select a quick and popular holdout estimation for feature and method selection purposes, with 20 varied runs. This holdout strategy randomly divides the training set of 40,403 phone contact into a practise and validation set with 2/3 and 1/3 of the phone contacts, respectively. Thereafter, the final results are collected by taking the average of the 20 runs. To test the statistical significance, we apply the non-parametric Mann-Whitney-Wilcoxon test at 5% confidence level. Second, we apply a rolling window estimation, which acts more like a realistic situation. The classification methods should be able to handle new phone contact data, otherwise we cannot make good predictions for the success of telemarketing and support managerial decision making. Therefore, we apply a rolling window scheme which updates new data and removes the oldest. Here, we use the most recent 20,000 phone contacts of the training set as the start of the training window, thereafter we renew the training window by substituting the oldest 10 phone contacts with the first 10 new phone contacts from the test set, and so forth. In total, and consequently, we apply 87 iterations to predict the full test set of 868 phone contacts. Finally, we compute AUC and ALIFT to present the method performance, where the method is better if the AUC and ALIFT are closer to 1.0. In addition, we include the running time of each method. As two methods could have almost equal metric results, but one of them has a faster calculation, this may influence the method preference.

4.1.1 Decision Tree

The decision tree is a popular data mining technique which produces a set of distinguishing values into a tree-like structure (Breiman, Friedman, Stone, & Olshen, 1984). It applies a Classification And Regression Tree (CART) algorithm that can deal with mixed categorical and numerical input variables. Here, $P(y_i \mid x_i)$ is the proportion of y_i categories over all elements of the leaf node that consists of input features x_i. The primary detriment of decision trees is the greedy construction which might cause overfitting. Yet, this results in a fast running time and comfortably understanding (Hastie, Tibshirani, & Friedman, 2009). In the study of Moro et al. (2014), the decision tree obtains the lowest AUC and ALIFT during the holdout technique. Therefore, we use the decision tree method as a benchmark to compare the other classification methods.

4.1.2 Logistic Regression

Another classification technique that is simple to understand is the logistic regression, because of the linear combination of its explanatory variables. This favored method elaborates a flatten nonlinear logistic adjustment over a multiple regression method and permits the estimation of class probabilities (Venables & Ripley, 2013):

$$P(y_i \mid x_i) = \frac{1}{1 + e^{\upsilon_0 + \sum_{f=1}^{G} \upsilon_f, x_{i,f}}}, \tag{1}$$

where $P(y_i \mid x_i)$ indicates the probability of class y_i with input features $x_i = (x_{i,1}, ..., x_{i,G})$ and υ_f indicates the parameters, set by the efficient BFGS algorithm (Moller, 1993). Nevertheless, the logistic regression is not suitable for complicated nonlinear situations and, according to Moro et al. (2014), this method performs worse in the rolling window prediction for predicting long-term bank deposits.

4.1.3 Support Vector Machine

A great advantage of the support vector machine is that this method can handle varying degrees of nonlinearity problems, in contrast to the previous two methods. The support vector machine applies a nonlinear Gaussian kernel: $K(x_j, x_l) = e^{-\gamma||x_j - x_l||^2}$, $\gamma > 0$ for different x_j and x_l (Hastie et al., 2009), and adopts the sequential minimal optimization through the 'kernlab' package (Platt, 1998). This algorithm converts the set of input features in a t-dimensional feature space and detects the best linear separating hyperplane in the feature space associated to support vector points. The result is collected as follows: $f(x_i) = \sum_{s=1}^{t} y_s \alpha_s K(x_s, x_i) + b$ and $p_i = \frac{1}{1+e^{Af(x_i)+B}}$, here $y_i \in \{0, 1\}$, α_s and b are the coefficients, t is the number of vectors and A and B are chosen through a regularized maximum likelihood problem (Wu, Lin, & Weng, 2004). However, this follows into a long running time and makes it difficult to understand (Dreiseitl & Ohno-Machado, 2002). To improve the performance of the support vector machine, the input features are standardized to a zero mean and one standard deviation (Hastie et al., 2009). Furthermore, we implement a grid search for the parameter of the kernel, σ, to detect the best hyperparameter, where σ has a range of 2^k with $k \in$ {-15, -13.2, -11.4, -9.6, -7.8, -6.0, -4.2, -2.4, -0.6, 1.2, 3.0} (Moro, Cortez, & Rita, 2014). In addition, the complex penalty parameter, C, is set to 3 (Cortez, 2010).

4.1.4 Artificial Neural Network

The artificial neural network is the best performing method in the study of Moro et al. (2014). This multilayer perceptron structure contains one hidden layer of hidden nodes and one output node. It is capable of preparing and training data such that the mean squared prediction error reduces to the minimum (Karouni, Daya, & Bahlak, 2011). Here, the state of the j-th neuron is calculated as follows: $s_j = f(w_{j,0} + \sum_{l \in P_j} w_{j,l} \times s_l)$, where f is the logistic function, $w_{j,l}$ denotes the weight between nodes j and l, P_j is the set of nodes reaching node j and $s_1 = x_{i,1}, ..., s_M = x_{i,M}$ for the input features $x_i = (x_{i,1}, ..., x_{i,M})$. According to Beucher, Møller and Greve (2019) the artificial neural network can operate with nonlinear problems, can deal with a large number of observations, prevents overfitting and is not sensitive to outliers. Therefore, this technique might be useful to predict the the success of telemarketing calls. To improve the artificial neural network, the input features are standardized to a zero mean and one standard deviation (Hastie et al., 2009) and we apply grid search for the size, $H \in \{0,1,2,3,5,6,7,8,9,10\}$. Moreover, the BFGS algorithm sets the number of distinct networks to 7 and the number of passes, *epochs*, is set to 100. (Moller, 1993).

4.1.5 Random Forest

In addition to the four methods from the study of Moro et al. (2014), we introduce the random forest to predict the success of telemarketing calls. We assume that tree algorithm methods might be good techniques for a rolling window scheme. Here, the random forest creates a forest-set of multiple decision trees, which are trained on randomly selected subsets. The outcome results from the average of the forest-set. The advantage of this technique is that it corrects for overfitting of a single decision tree and is still straightforward to understand (Hall & Holmes, 2003). To improve the random forest, the input features are standardized to a zero mean and one standard deviation, just as the support vector machine and the artificial neural network (Hastie et al., 2009). Moreover, we apply a grid search for the following 4 parameters: the number of trees, $ntree \in \{100, 500, 1000\}$; the minimum number of samples required to split each node, $min\_samples\_split \in \{2, 4, 6\}$; the maximum depth of each tree, $max\_depth \in \{1, 3, 5, 7, 9, 11\}$; and the minimum number of samples required to be at a leaf node, $min\_samples\_leaf \in \{1, 3, 5, 7\}$.

4.2 Contact Time Prediction

We apply the logarithmic transformed ordinary least squares regression and two count data regressions: the poisson regression and the negative-binomial regression, to predict the contact time of telemarketing calls for selling long-term bank deposits. An explanation of those methods, including their advantages and disadvantages, is given below. All contact time code is programmed in R with the package 'MASS' (Ripley et al., 2013). To select the most appropriate method, we use the mean squared prediction error of the test set based on the training set:

$$MSPE = (contact.time_i - predict(contact.time_i))^2 , \tag{2}$$

where the $contact.time_i$ is the (self-created) total contact time of a campaign. The average contact time of a campaign, based on the training set, equals 612.5 seconds. To realise the performance of the three regression methods, we include the average benchmark (ABM). This benchmark assumes that the $predict(contact.time_i)$ equals 612.5 seconds for every potential customer call. If one of the regression methods has a lower mean squared prediction error than the benchmark, we can, at least, provide a better prediction than the average.

4.2.1 Logarithmic Transformed Ordinary Least Squares Regression

We apply a logarithmic transformed ordinary least squares regression to predict the total contact time,

$$\log(contact.time_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_p x_{i,p} + \epsilon. \tag{3}$$

Here, the β's are estimated values for the input features, p indicates the number of input features and ϵ is the error term. As the distribution of the contact time is positively skewed, the logarithmic transformation of the dependent time variable makes it more like the normal distribution, which is valuable for the assumptions of inferential statistics (Changyong et al., 2014). Through the studentized Breusch-Pagan test we discover heteroskedasticity, more details are shown in Appendix C (1). This indicates that the spread of the residuals over the range of measured values is not constant. Since the ordinary least squares regression assumes that all residuals have a constant variance, we correct this through White standard errors (Heij et al., 2004, page 325) .

Without taking the logarithm of the total contact time, the dependent time variable is a discrete ordered value and represents a quantity. The goal is to examine the correlation between the count dependent variable *contact.time* and the 21 input features. Two methods, which might be suitable for this examination are the poisson regression and the negative-binomial regression, since the dependent contact time variable has a non-normal distribution and non-negative, discrete and ordered values which are close to zero (Winkelmann, 1995).

4.2.2 Poisson Regression

We assume that the $contact.time_i$ is poisson distributed, that is, $contact.time_i \sim \mathrm{POI}(\mu)$:

$$Pr[contact.time_i = l] = \frac{e^{-\mu}\mu^l}{l!}, \tag{4}$$

for $\mu > 0$ and l is the contact time in seconds. This generalized linear method is a flexible generalization of ordinary least squares regression that allows response variables to have error distribution models other than a normal distribution. The mean and variance of $contact.time_i$ are equal to μ:

$$E[contact.time_i] = \mu, \quad V[contact.time_i] = \mu. \tag{5}$$

However, μ is unknown and a μ constant across individuals is implausible in (5). Therefore, we make μ dependent on the 21 input features x:

$$\mu = e^{\beta' x}. \tag{6}$$

Here, (4) and (6) result in the following:

$$Pr[contact.time_i = l|x] = \frac{e^{-(e^{\beta' x})}(e^{\beta' x})^j}{j!}, \tag{7}$$

with

$$E[contact.time_i|x] = e^{\beta' x}, \quad V[contact.time_i|x] = e^{\beta' x}. \tag{8}$$

The detriment of the poisson regression is that the variance equals the mean. In practice, data often show overdispersion (Dean & Lawless, 1989), that is:

$$E[contact.time_i|x] < V[contact.time_i|x]. \tag{9}$$

As we find overdispersion through testing $H_0 = V[contact.time_i|x] = E[contact.time_i|x] = \mu$, we apply another generalized linear method: the negative-binomial distribution. The overdispersion test details are shown in Appendix C (2).

4.2.3 Negative-Binomial Regression

Here we assume that the $contact.time_i$ is negative-binomial distributed, that is, $contact.time_i \sim$ NB-2(α,μ):

$$Pr[contact.time_i = l|x] = \frac{\Gamma(l + \alpha^{-1})}{\Gamma(l+1)\Gamma(\alpha^{-1})} \left(\frac{e^{\beta' x}}{e^{\beta' x} + \alpha^{-1}}\right)^l \left(\frac{\alpha^{-1}}{e^{\beta' x} + \alpha^{-1}}\right)^{\frac{1}{\alpha}}, \tag{10}$$

for $\mu > 0$, $\alpha > 0$, x are the 21 input features and l is the contact time in seconds. The mean and variance of $contact.time_i$ are given as follow:

$$E[contact.time_i|x] = e^{\beta' x}, \quad V[contact.time_i|x] = e^{\beta' x}(1 + e^{\beta' x}\alpha). \tag{11}$$

Accordingly, for $\alpha > 0$ the is variance greater than the mean.

4.3 Characteristics and the best set of potential costumers

4.3.1 Success and contact time characteristics

We apply the method with the highest AUC and ALIFT to determine which characteristics influence the success of telemarketing calls. For this purpose, we fit the training data set and analyze the response when a given input feature is varied through its domain by mediation of the data-based sensitivity analysis algorithm (Cortez & Embrechts, 2013). Then, we visualize the results in an importance bar plot. Moreover, we plot the 4 major characteristics through the variable effect characteristic (VEC) curve. This permits us to understand the overall influence of an input feature in the predicted outcome by plotting the input feature range of values against the average sensitivity responses. The input values on the horizontal axis of the VEC curve are scaled, because of the various values for each input feature. By doing this, we are able to compare the different input features in one plot.

To determine which characteristics influence the time of telemarketing calls, we apply the method with the lowest mean squared prediction error and fit the training set. Due to the fact that we use regression methods which result in predicted β's and standard errors, we detect the contact time characteristics through the significance and the sign of the input features.

4.3.2 The best set of potential costumers

We create a self-made function to obtain the $Customer\_Performance$:

$$Customer\_Performance = \frac{probability\_of\_success}{(contact.time + 180)}. \tag{12}$$

Here, the $probability\_of\_success$ is the predicted probability of success from the best classification method in the rolling window estimation, where $probability\_of\_success \in [0,1]$. The $contact.time$

is the predicted contact time of a call determined by best regression technique, where *contact.time* $\in [0, \infty)$. The $Customer\_Performance$ is a ratio of success probability per second. We rank this from high to low to obtain the best set of potential costumers.

To create a more realistic situation, we add 180 seconds to the total time of a phone contact. Since the call center is not able to make calls every second of the day, because of the breaks between the phone contacts and the time it takes to find information of the next potential costumer, we assume an extra time of 3 minutes should be added for a more businesslike setting. Furthermore, without the extra 180 seconds, a potential costumer with a success rate of 10% and a total contact time of 100 seconds gets a higher Customer_Performance than a potential costumer with and success rate of 90% and a total contact time of 1,000 seconds, which should not be plausible. Therefore, the three added minutes also create a better best set of potential costumers.

5 Results

5.1 Success results

Table 3 shows the results of the holdout estimation to predict the success of bank telemarketing calls. The best obtained parameters are shown in parentheses. Based on the AUC and ALIFT, the artificial neural network with a size of 3 and a number of distinct networks of 7, performs the best with respectively 0.781 and 0.752. However, the logistic regression (AUC = 0.780, ALIFT = 0.751) and the random forest (AUC = 0.769, ALIFT = 0.740) perform almost as good as the artificial neural network for both metrics and the logistic regression even has less running time. The benchmark decision tree performs the worst and is the fastest method with an AUC of 0.682, an ALIFT of 0.597 and a running time of 29.2 seconds, which was expected because of greedy tree construction.

Table 3: Results success prediction for holdout estimation

metric	DT	LR	SVM$^{(y=2^{-0.6},C=3)}$	ANN$^{(H=3,Nr=7)}$	RF$^{(nt=1000,msp=4,md=3,msl=1)}$
AUC	0.682	0.780	0.712	0.781	0.769
ALIFT	0.597	0.751	0.689	0.752	0.740
Running Time	29.2	42.7	8734.2	441.7	3377.1

Notes. The results of the holdout estimation to predict the success of bank telemarketing calls are shown. The area under the receiver operating characteristic curve and area under the LIFT cumulative curve are used to compare the method performance. The dependent variable is dummy y, which specifies whether the call results into a deposit sale. Furthermore, the input features contain personal, contact, historic and economic information. The values in parentheses are the best resulting parameters through the grid search. The running time is in seconds.

In Table 4 the results of the rolling window estimation are shown. The best produced parameters are shown in parentheses. Here, we obtain a more realistic environment. The artificial neural network still generates the best results, with a running time of 1644.3 seconds, an AUC of 0.757 and an ALIFT of 0.632. Furthermore, the results of the logistic regression (AUC = 0.742, ALIFT = 0.625) have decreased more than the random forest (AUC = 0.742, ALIFT = 0.626), when we compare the holdout estimation results with the rolling window results. Therefore, we conclude that the tree algorithm random forest can handle more realistic situations. However, the artificial neural network is the best method to predict the success of telemarketing calls for both the holdout and rolling window estimation.

Table 4: Results success prediction for rolling window estimation

metric	DT	LR	SVM$^{(y=2^{-0.6},C=3)}$	ANN$^{(H=3,Nr=7)}$	RF$^{(nt=1000,msp=4,md=3,msl=1)}$
AUC	0.652	0.742	0.668	0.757	0.745
ALIFT	0.579	0.625	0.586	0.632	0.626
Running Time	53.9	89.2	20514.6	1644.3	10106.3

Notes. The results of the rolling window estimation to predict the success of bank telemarketing calls are shown. The area under the receiver operating characteristic curve and area under the LIFT cumulative curve are used to compare the method performance. The dependent variable is dummy y, which specifies whether the call results into a deposit sale. Furthermore, the input features contain personal, contact, historic and economic information. The values in parentheses are the best resulting parameters through the grid search. The running time is in seconds.

5.2 Contact time results

Table 5 shows the results of the contact time prediction methods. All three regression techniques obtain a lower mean squared prediction error than the average benchmark (MSPE = 1,003,509), which means we can provide a better prediction than the average of 612.5 seconds. The negative-binomial regression obtains the lowest mean squared prediction error of 551,313. The poisson regression (MSPE = 551,374) performs worse than the negative-binomial regression. Despite the logarithmic transformation to correct for skewness, the logarithmic transformed ordinary least squares regression produces the highest mean squared prediction error. Thus, we conclude that the negative-binomial regression is the best method to predict the contact time of telemarketing calls.

Table 5: Results contact time prediction

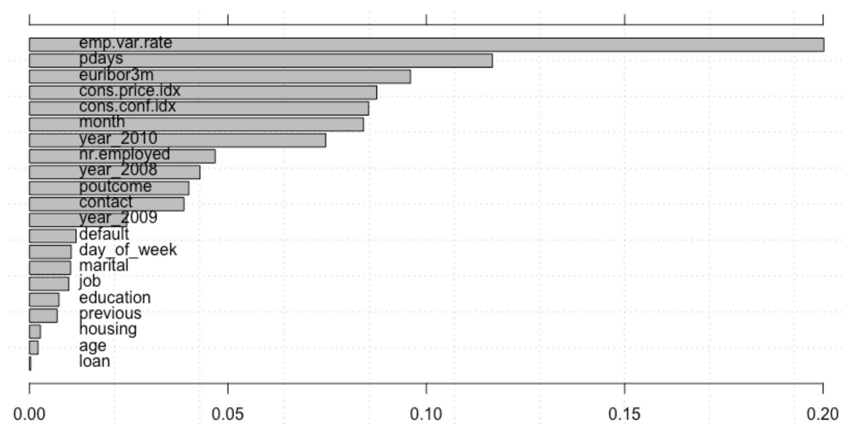
metric	ABM	OLS	POI	NBM
MSPE	1,003,509	636,401	551,374	551,313

Notes. The mean squared prediction error of the benchmark and the contact time regressions are shown. The benchmark assumes that the predicted contact time equals 612.5 seconds, the average contact time of a customer based on the training set, for every potential customer. We apply the logarithmic transformed ordinary least squares regression, the poisson regression and the negative-binomial regression. The dependent variable is the logarithm of the contact time and the input features contain personal, contact, historic and economic information.

5.3 Characteristics results and the best set of potential costumers

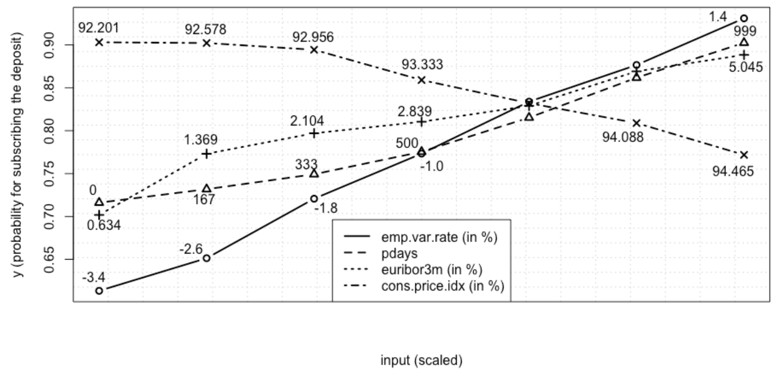
We apply the best performing classification method, the artificial neural network, to determine which characteristics influence the success of telemarketing calls. In Figure 1, we visualize the results of the data-based sensitivity analysis in an importance bar. The most influential input feature is the quarterly variation of the employment rate with 20.02%, followed by the number of days since the last call for any other campaign with 11.67%. It appears that 5 out of 8 most important characteristics are economic input features. The weakest input feature is the dummy variable which specifies whether the potential customer has a personal credit, this loan features has an importance of 0.21%. Furthermore, the three 'year' variables are located in the 12 most important characteristics. Therefore, we conclude that the addition of these variables is advantageous.

Figure 1: Importance bar



The four major success characteristics (emp.var.rate, pdays, euribor3m and cons.price.idx) are shown through the variable effect characteristic curve in Figure 2. The employment variation rate, the number of days since the last call for any other campaign and the Euribor rate have an upward trend. This means that an increase in input leads to an increase in the success probability. The consumer price index has the opposite effect, an increase leads to an decrease in the success probability.

Figure 2: VEC curve of 4 major characteristic



We apply the negative-binomial regression to determine which characteristics influence the contact time of telemarketing calls. The results of the significant coefficients are shown in Table 6. The full regression results, including insignificant coefficients, can be found in Appendix D. Note that the coefficient indicates that a one-unit increase depends on the expected logarithm of the contact time. We find that potential customers with loans in delay and an increase in the Euribor rate have the largest significant negative effects on the contact time. Since an increase in the Euribor rate also strongly increases the probability of success, this input feature is for both the success rate and the contact time an important variable for efficiency.

Table 6: negative-binomial regression results

	coefficient
jobself-employed	0.105 (0.029) ***
jobstudent	0.074 (0.040) *
jobunknown	-0.122 (0.058) **
maritalunknown	0.429 (0.114) ***
educationbasic.9y	-0.054 (0.021) **
educationhigh.school	-0.038 (0.022) *
educationprofessional.course	-0.044 (0.025) *
defaultyes	-1.201 (0.583) **
housingunknown	-0.063 (0.033) *
contacttelephone	0.070 (0.020) ***
monthaug	-0.157 (0.070) **
monthdec	0.395 (0.103) ***
monthjul	0.239 (0.032) ***
day_of_weekmon	-0.047 (0.016) ***
day_of_weekthu	-0.033 (0.016) **
day_of_weektue	-0.044 (0.016) ***
day_of_weekwed	-0.028 (0.016) *
pdays	-0.0003 (0.0001) ***
poutcomenonexistent	0.112 (0.039) ***
emp.var.rate	0.347 (0.063) ***
cons.conf.idx	0.019 (0.005) ***
euribor3m	-0.205 (0.092) **

Notes. The negative-binomial regression results are shown. The dependent variable is the logarithm of the contact time and the input features contain personal, contact, historic and economic information. We only present the significant coefficients, since those values can be interpreted, and the standard errors are given in parentheses. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

We use the predicted success rate from the artificial neural network in the rolling window estimation and the contact time predictions from the negative-binomial regression in a self-made function. Here, we select the best set of 868 potential customers based on the highest success probability per second. In Table 7, the 10 best and 2 worst potential customers are shown. Through this, we advice the bank to call potential customer 210 first, because he or she has the highest customer performance of 0.1213. Important to mention is that, for example, potential customer 727, rank 4, has a lower probability of success than potential customer 73, rank 5. However, because of the contact time, we find that potential customer 727 has a higher success probability per second. Therefore, potential customer 727 is more interesting for the bank and we advice to call him/her before potential customer 73. Potential costumer 397, rank 868, is at the bottom of the list. With a predicted success probability of 4.51% and a contact time of 982.41 seconds, he or she has the lowest success probability per second. Therefore, we advice the bank to call potential costumer 397 the latest.

Table 7: Best set of potential customers

rank	potential customer	success probability	contact.time	Customer Performance
1	210	76.33	629.05	0.1213
2	75	74.69	622.85	0.1199
3	77	72.41	621.87	0.1164
4	727	70.74	624.60	0.1133
5	73	75.14	666.07	0.1128
6	78	68.13	606.21	0.1124
7	192	67.86	608.85	0.1115
8	737	71.19	641.91	0.1109
9	2	65.79	595.71	0.1104
10	72	74.61	679.10	0.1099
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
867	400	7.53	847.23	0.0089
868	397	4.51	982.41	0.0046

Notes. The best set of 868 potential customers are shown based on the Customers Performance. The rank has a range from 1 to 868. The potential customer number indicates a potential customer in the test set. The success probability indicates the change of a sale based on the artificial neural network in the rolling window estimation. The predicted contact time comes from the negative-binomial regression. Finally, the Customer Performance is the success probability per second.

6 Conclusion and Discussion

6.1 Conclusion

In this study, we predict the success rate and contact time of telemarketing calls for selling long-term bank deposits. Furthermore, we determine the influence of potential customer characteristics and present the best set of potential customers. Therefore, we pose the following research questions:

- *To what extent are various classification methods capable of examining the success of telemarketing calls for selling long-term bank deposits?*
- *And, to what extent are regression methods capable of forecasting the contact time of telemarketing calls for selling long-term bank deposits?*
- *And, which characteristics influence the success rate and contact time of telemarketing calls for selling long-term bank deposits?*
- *Finally, what is the best set of potential customers for selling long-term bank deposits?*

Moro et al. (2014), Jaing (2018) and Selma (2020) have applied multiple data-driven methods to improve the success of telemarketing calls for selling long-term bank deposit and they obtained successful results. However, we did find a few areas for improvement. First, they applied input features which are only known after the call and that is not practicable for a marketeer to apply in a prediction method. Furthermore, the tree algorithm methods are not optimally used. Lastly, the contact time has not been taken into account. An attractive way to advice companies is by finding a way to reduce the contact time while the success rate remains the same. Therefore, we first predict the success rate of telemarketing calls through five classification methods: the decision tree, the logistic regression, the support vector machine, the artificial neural network and the random forest. Then, we predict the contact time of telemarketing calls through three regression techniques: the transformed ordinary least squares regression, the poisson regression and the negative-binomial regression.

We apply the best performing classification method, the artificial neural network (AUC = 0.757, ALIFT = 0.632), and the best performing regression, the negative-binomial regression (MSE = 551,313), to determine which characteristics influence the success rate and contact time. The employment rate, the number of days since the last call for any other campaign and the Euribor rate are most influential for the success rate. Subsequently, we detect that potential customers with loans in delay and an increase in the Euribor rate have the largest significant negative effects on the contact time.

To find the best set of the most recent 868 potential customers, we use the predicted success rate from the artificial neural network and the contact time forecasts from the negative-binomial regression in a self-made function. Based on the highest success probability per second, we are able to create a list, where the first potential customer on the list is most attractive to call.

6.2 Discussion and Outlook

To improve our study, we propose further research for the limitations we faced. First, we made two dominant assumptions. The dependent variable *contact.time* is created by multiplying the duration time of the most recent call and the number of contacts for the same campaign. Since not all the calls have the same duration, this self-created variable is not equivalent to actual contact time. Therefore, we suggest to apply the actual contact time in further research to make the study more realistic. Second, to select the best set of potential customers, we assume that an extra 3 minutes should be added to the contact time for a more realistic setting. We include this because of the short breaks between the phone contacts. However, as this is a new implementation, it requires further research to determine the accuracy. Therefore, we advise to focus more on the realistic time breaks in further research. Third, in their study, Moro et al. (2014) analyzed 150 different input features. We have received a data set of only 21 input features, so there might be some important input features that are not included in our study. We propose to obtain the data set of 150 features in the future. This might result in better use of input features and an increase in metric performances. Fourth, we do not focus on the correlation between the 21 input features. For example, the number of employees in Portugal and the employment variation rate are highly correlated, which results in prediction problems. Therefore, we advise to take correlation into account and try to prevent multicollinearity. Finally, the number of days since the last call for any other campaign is indicated with '999' days if there was no previous call. In further research we suggest to correct this.

References

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues*, *9*(5), 272.

Beucher, A., Møller, A. B., & Greve, M. H. (2019). Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark. *Geoderma*, *352*, 351–359.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees.* USA, Ohio: CRC press.

Changyong, F., Hongyue, W., Naiji, L., Tian, C., Hua, H., Ying, L., et al. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105.

Chen, M.-S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, *8*(6), 866–883.

Cortez, P. (2010). Data mining with neural networks and support vector machines using the R/rminer tool. In *Industrial Conference on Data Mining* (pp. 572–583).

Cortez, P. (2020). Package 'rminer'. *Teaching Report*, 59.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modelling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, *47*(4), 547–553.

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, *225*, 1–17.

Dean, C., & Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, *84*(406), 467–472.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 498–506.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, *35*(5-6), 352–359.

Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering*, *15*(6), 1437–1447.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* DE, Berlin: Springer Science & Business Media.

Heij, C., de Boer, P., Franses, P. H., Kloek, T., van Dijk, H. K., et al. (2004). *Econometric Methods with Applications in Business and Economics.* UK, Oxford: Oxford University Press.

Jiang, Y. (2018). Using logistic regression model to predict the success of bank telemarketing. *International Journal on Data Science and Technology*, *4*(1), 35.

Karouni, A., Daya, B., & Bahlak, S. (2011). Offline signature recognition using neural networks approach. *Procedia Computer Science*, *3*, 155–161.

Moller, A. (1993). A scaled conjugate gradient algorithm for fast supervised learning neural networks, 6,(1993), 525–533. *Google Scholar Digital Library*.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22–31.

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, *52*(2), 464–473.

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Technical Report MSR-TR-98-14, Microsoft Research*.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'MASS'. *Cran R*, *538*.

Selma, M. (2020). Predicting the success of bank telemarketing using artificial neural network. *International Journal of Economics and Management Engineering*, *14*(1), 1–4.

Venables, W. N., & Ripley, B. D. (2013). *Modern Applied Statistics with S-PLUS*. DE, Berlin: Springer Science & Business Media.

Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. *Journal of Business and Economic Statistics*, *13*(4), 467–474.

Wu, T.-F., Lin, C.-J., & Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, *5*(Aug), 975–1005.

A Appendix

Table 8: Detailed analyses numerical input features

contact time		age		pdays		previous		
	min	0.0	min	17.0	min	0.0	min	0.0
	1st Qu.	165.0	1st Qu.	32.0	1st Qu.	999.0	1st Qu.	0.0
	mean	611.8	mean	40.0	mean	962.5	mean	0.2
	3rd Qu.	669.0	3rd Qu.	47.0	3rd Qu.	999.0	3rd Qu.	0.0
	max	30186	max	98.0	max	999.0	max	7.0
emp.var.rate		cons.price.idx		cons.conf.idx		euribor3m		
	min	-3.4	min	92.2	min	-50.8	min	0.6
	1st Qu.	-1.8	1st Qu.	93.1	1st Qu.	-42.7	1st Qu.	1.3
	mean	0.1	mean	93.6	mean	-40.5	mean	3.6
	3rd Qu.	1.4	3rd Qu.	94.0	3rd Qu.	-36.4	3rd Qu.	5.0
	max	1.4	max	94.8	max	-26.9	max	5.0
nr. employed		campaign		duration				
	min	4964.0	min	1.0	min	0.0		
	1st Qu.	5099.0	1st Qu.	1.0	1st Qu.	102.0		
	mean	5167.0	mean	2.6	mean	258.3		
	3rd Qu.	5228.0	3rd Qu.	3.0	3rd Qu.	319.0		
	max	5228.0	max	56.0	max	4918.0		

Notes. A detailed analyses of all numerical input features is shown. The minimum, the first quartile, the mean, the third quartile and the maximum are given.

Table 9: Detailed analyses categorical input features

y		job		marital		education		
	no	36537	admin.	10419	divorced	4611	university.degree	12164
	yes	4639	blue-collar	9253	married	24921	high.school	9512
			technician	6739	single	11564	basic.9y	6045
			services	3967	unknown	80	professional.course	5240
			management	2924			basic.4y	4176
			retired	1718			basic.6y	2291
			other	6156			other	1748
default		housing		loan		contact		
	no	32577	no	18615	no	33939	cellular	26135
	yes	8596	yes	21571	yes	6248	telephone	15041
	unknown	3	unknown	990	unknown	990		
month		day		poutcome				
	may	13767	mon	8512	failure	4252		
	jul	7169	tue	8086	nonexistent	35551		
	aug	6176	wed	8134	success	1273		
	jun	5318	thu	8618				
	nov	4100	fri	7826				
	apr	2631						
	other	2015						

Notes. A detailed analyses of all numerical input features.

B Appendix

(a) Contact time

(b) Log(Contact time)

Figure 3: Histograms of the contact time

C Appendix

(1)

Studentized Breusch-Pagan test:

BP = 800.63

df = 52

p-value $< 2.2 \times 10^{-16}$

(2)

Overdispersion test:

z = 21.488

p-value $< 2.2 \times 10^{-16}$

sample estimates dispersion = 1509.768

D Appendix

	coefficient
(Intercept)	-2.16 (13.72)
age	0.00 (0.00)
jobblue-collar	0.01 (0.02)
jobentrepreneur	0.02 (0.03)
jobhousemaid	-0.04 (0.03)
jobmanagement	-0.01 (0.02)
jobretired	0.02 (0.03)
jobself-employed	0.10 (0.03) ***
jobservices	0.01 (0.02)
jobstudent	0.07 (0.04) *
jobtechnician	-0.0 (0.02)
jobunemployed	0.01 (0.03)
jobunknown	-0.12 (0.06) **
maritalmarried	0.01 (0.02)
maritalsingle	0.01 (0.02)
maritalunknown	0.43 (0.11) ***
educationbasic.6y	-0.04 (0.03)
educationbasic.9y	-0.05 (0.02) **
educationhigh.school	-0.04 (0.02) *
educationilliterate	-0.10 (0.24)
educationprofessional.course	-0.04 (0.02) *
educationuniversity.degree	-0.02 (0.02)
educationunknown	-0.04 (0.03)
defaultunknown	-0.02 (0.01)
defaultyes	-1.20 (0.58) **
housingunknown	-0.06 (0.03) *
housingyes	-0.02 (0.01)
loanyes	0.01 (0.01)
contacttelephone	0.07 (0.02) ***
monthaug	-0.16 (0.07) **
monthdec	0.39 (0.10) ***
monthjul	0.24 (0.03) ***
monthjun	0.00 (0.05)
monthmar	-0.07 (0.06)
monthmay	-0.03 (0.03)
monthnov	0.09 (0.08)
monthoct	-0.08 (0.07)
monthsep	0.03 (0.09)
day_of_weekmon	-0.05 (0.02) ***
day_of_weekthu	-0.03 (0.02) **
day_of_weektue	-0.04 (0.02) ***
day_of_weekwed	-0.03 (0.02) *
pdays	-0.00 (0.00) ***
previous	0.00 (0.03)
poutcomenonexistent	0.11 (0.04) ***
poutcomesuccess	-0.17 (0.11)
emp.var.rate	0.35 (0.06) ***
cons.price.idx	0.03 (0.09)
cons.conf.idx	0.02 (0.01) ***
euribor3m	-0.20 (0.09) ***
nr.employed	0.00 (0.00)
year_2008	-0.39 (0.49)
year_2009	0.16 (0.15)

Notes. The negative-binomial regression results are shown. The dependent variable is the logarithm of the contact time and the input features contain personal, contact, historic and economic information and the standard errors are given in parentheses. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.