Predicting the Success of Bank Telemarketing using Neural Networks and Random Forest Algorithms

1st Brendan Moran
School of Computing
National College of Ireland
Dublin, Ireland
x15024849@student.ncirl.ie

***Abstract*—In this paper the author examines the data of a large scale bank marketing exercise of over 41,000 individual phone calls to sell a term deposit to customers. The purpose of the exercise is to improve the bank's call/success ratio by identifying customers who have a low probability of taking up the offer and not including them in the the to be called list in future campaigns. Classifying customers into call/don't call groups and producing a very small call list is not the objective but rather reducing the call list and retaining all the successful outcomes in the reduced list. Because of the sophisticated nature of the data and the classification problem, simplistic models such as K-Nearest Neighbour and Decision trees were not suitable to achieve the aim stated above. More complex models such as Random Forest and Neural Network were more appropriate to tackle this classification problem as there is no dependent variable in the data set ie. no metric which shows what pool of customers is best to call apart from successful calls from the previous campaign. The results from the following are that Random Forests predict well in a smaller data set while Neural Networks predict well in a larger one (80% of the data set)**

***Index Terms*—Bank telemarketing, Random Forest, Neural Networks**

I. Introduction

Sales and Marketing companies usually employ a typical strategy to enhance business. This is usually in the form of direct target marketing. This is usually carried out in a centralised contact centre or call centre. Agents contact customers through various means of communication such as phone, email and direct message chat. For Sales and Marketing phone is the preferred choice since agents can usually influence customers a lot easier over the phone. Contacts are divided into two categories, outbound and inbound calls with outbound calls naturally being the more intrusive since the agent is contacting the customer. Outbound calls can further be subdivided into cold and warm calls. If the contact has never been in touch with the company before then this is deemed a cold call. Technology enables a new shift in marketing as there can be a focus on maximizing customer success by using all available data and metrics to build a tailored customer experience which align more closely with their wishes. In the past agents would have to call every customer on a contact list. Selecting the best set of clients which are more likely to subscribe.[7]The smart selection of customers, using predictive modeling, thus is of crucial importance. Done well, it can result in substantial additional profits compared with random selections of customers for targeted retention campaigns [1]Data mining is defined as a process that uses mathematical, statistical, artificial intelligence and machine learning techniques to extract and identify useful information and subsequently gain knowledge from databases. [4] Describes Data Mining as The automated extraction of predictive information from (large) databases. The most important words in that sentence are automated and predictive, data mining lets you be proactive rather than reactive, prospective rather than being retrospective. [3] Common data mining practice in classification is to gather a great number of variables and apply different standard algorithms. Given the set of predefined classes and a number of attributes, these classification methods can provide a model to predict the class of other unclassified data. Mathematical techniques that are often used to construct classification methods are binary decision trees, neural networks, and logistic regression. **Figure 1.** shows a typical data mining framework.

In this study the author will be using data mining in order to reduce the number of calls made while not reducing the number of successful calls. Calling customers who share similar characteristics to the customers who previously bought a term deposit is the aim. This classification problem is still supervised learning as the answer is still available, but the author is looking for a variant of the answer available in the data set. There are two measures of success, what percentage of the successful calls did we keep and how well did we eliminate unsuccessful calls while doing it. With this in mind simplistic data mining models such K Nearest Neighbour (KNN) and Decision Trees are too simplistic for a classification problem of this nature.

II. Literature Review

The data which is used in this work is the data of customers of a Portuguese banking institution. The most relevant study was [1] which aimed to find the model that can increase the success rate of telemarketing for the bank. The used a variety of data statistical techniques such as Support Vector Machine (SVM), Decision Tree(DT) and Naive Bayes. According to their results SVM comes up with most efficient results. With regard to features Call Duration was the most relevant feature which states that the longer a call lasts the greater chance of success. In the authors study this metric was removed as it is not available in the data set chosen so it was dropped.

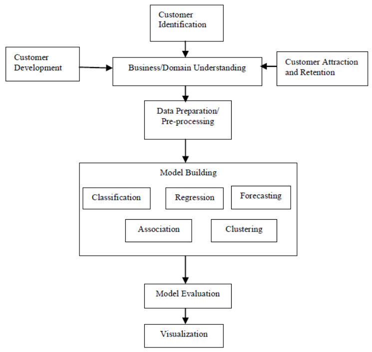


Fig. 1. Data Mining Framework

[2] Tried to predict the success of a Telemarketing campaign from data collected from 2008 to 2013. A large data set with 150 features was reduced to 22 features and applied four data mining models Logistic Regression, decision trees, support vector machine and neural networks(NN). NN performed the best while decision trees discloses that probability of success in inbound calls are greater. [4] Uses the same data set as per this study but only uses 10% of the pre-processed data set is used. A Multi-Layer perception neural Network using back propagation is used.This "helps in learning, and reducing the future errors5. Good learning ability, fast real-time operation, less memory demand, analysis of complex patterns are some of the advantages of MLPNN and the disadvantages include high-quality data requirement of the network, careful selection of variables etc." [10] Favours a simplistic regression model is just as viable as more advanced algorithms such as neural networks and random forests."They require additional parameter tuning and often fail to provide direct churn probabilities, so they demand additional,post-processing steps such as calibration." [12] Supervised learning requires input data that has both predictor (independent) attributes and a target (dependent) attribute whose value is to be estimated. In addition, the process learns how to model (predict) the value of the target attribute based on predictor attributes. The author uses three statistical measures; classification accuracy, sensitivity and specificity. False Negative (FN) **(see Figure 2)** is the number of incorrect predictions that an instance is false. Because of the nature of cold calling we expect a lot of false positives - it's cold calling - what else should we expect? - but false negatives (i.e deciding not to call a customer who would buy) is a significant problem. [5] Uses a new approach, the **RFM** method. R represents the period since the last purchase. F is the number of purchases made by a customer during a certain period. M is the total purchase amount by a customer over that period.The data set employed in this research has the information about who responded to the direct marketing or campaign.Using R, F, and M as three predictive variables, each data mining technique will develop a binary customer response model based on the training data set and apply the model to the test data set. This will generate prediction accuracy rate the percentage of customers classified correctly.

		Predicted	
		Positive (yes)	Negative (no)
Actual	Positive (yes)	TP	FP
	Negative (no)	TN	FN

Fig. 2. Confusion Matrix

III. Theory and Models

Neural Networks

[9] One of the greatest benefits of ANNs are their capacity to deal with datasets containing large number of observations. It can also manage to estimate the nonlinear relationship. [10] Artificial neural networks are directed machine learning approaches which regulate the relationship between the set which is known for the training points and distinctive natural attributes with intention of classification for the new one. It prevent the model from over fitting and also is not affect by the outliers. Artificial Neural Network resembles the human brain in learning over the data storage and training. It is made and prepared over a particular input information training pattern. Throughout the procedure of learning, the results of NN is then matched to the target value and in this way, algorithm has been accomplished to reduce the error as minimum between the two values.

Random Forest

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees. The ensemble used is known as bagging as successive trees do not depend on earlier trees, each is independently constructed using a different bootstrap sample of the dataset. Predictions are made by taking a simple majority vote. In decision trees each node is split using the best split among all variables, whereas in a random forest, each node is split using the best among a subset of predictors randomly chosen at that node. It was chosen as it provides an additional layer of randomness and therefore is quite robust against over-fitting. The parameters used in this dataset were the default parameters for Random Forest.

IV. Data Preprocessing

The source is : https://archive.ics.uci.edu/ml/machine-learning-databases/00222/.The data is a outcome of a large scale bank marketing exercise of over 41,000 individual phone calls to customers to sell a term deposit product. The data includes the outcome 'yes' or 'no' based on the customer taking out the term discount product within a set period after the campaign. There 21 independent variables in the dataset which are Age ,Job, Marital status, Education, Credit Default, Housing, Loan, Contact, Month etc. as well as several economic indicators **(see Fig 11.)** The data also includes the duration of the phone call - which correlates highly with the outcome and is not known before the customer is called so it is removed from the data as it is not relevant to the choice of a customer for a marketing call. The purpose of the exercise is to improve the bank's call/success ratio by identifying customers who have a low probability of taking up the offer and not including them in the customer to be called list in future campaigns.

	X	non.missing	missing	missing.percent	unique	mean	min	p1	p5	p10	p25	p50	p75	p90	p95	p99	max
1	age	41188	0	0	78	40.02	17.00	23.00	26.00	28.00	32.00	38.00	47.00	55.00	58.00	71.00	98.00
2	duration	41188	0	0	1544	258.29	0.00	11.00	36.00	59.00	102.00	180.00	319.00	551.00	752.65	1271.13	4918.00
3	campaign	41188	0	0	42	2.57	1.00	1.00	1.00	1.00	1.00	2.00	3.00	5.00	7.00	14.00	56.00
4	pdays	41188	0	0	27	962.48	0.00	3.00	999.00	999.00	999.00	999.00	999.00	999.00	999.00	999.00	999.00
5	previous	41188	0	0	8	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	2.00	7.00
6	emp.var.rate	41188	0	0	10	0.08	-3.40	-3.40	-2.90	-1.80	-1.80	1.10	1.40	1.40	1.40	1.40	1.40
7	cons.price.idx	41188	0	0	26	93.58	92.20	92.20	92.71	92.89	93.08	93.75	93.99	94.47	94.47	94.47	94.77
8	cons.conf.idx	41188	0	0	26	-40.50	-50.80	-49.50	-47.10	-46.20	-42.70	-41.80	-36.40	-36.10	-33.60	-26.90	-26.90
9	euribor3m	41188	0	0	316	3.62	0.63	0.66	0.80	1.05	1.34	4.86	4.96	4.96	4.97	4.97	5.04
10	nr.employed	41188	0	0	11	5167.04	4963.60	4963.60	5017.50	5076.20	5099.10	5191.00	5228.10	5228.10	5228.10	5228.10	5228.10

Fig. 3. Data Quality Report

A. Initial Data Treatment

The data arrives as semi-colon separated data - an unusual format as most data is usually in comma separated value format (CSV) The author chose to convert outside.Read file and convert while reading, then read.csv routine but apply it to data read with readLines and with gsub providing replacement of semicolons with commas. Text Connection makes the result of the gsub on readLines look like a file to enable us to use read.csv on it. The data includes some economic indicators which are the same for many rows these have no bearing on the inclusion of the customer in a call list so we will drop these from the analysis data.However this data could be analysed and used by the bank to determine when to increase its marketing activity and when to reduce it. If there are correlations between general economic indicators and the success of the campaign dropping 'economic' columns : emp.var.rate, cons.price.idx, cons.conf.idx, euribor 3m, nr.employed. Dropping these 5 variables would be a good course of action.We know that the call duration closely correlates to the outcome but it is not available before the customer is called so we need to drop it. Also the number of contacts performed during this campaign is not available as an initial selection indicator so we have to drop it as well. This leaves 14 variables. Asapply(bankFull, function(x) sum(is.na(x))) appears to show no missing values in the dataset. To confirm the presence or absence of missing values in the dataset, a data quality report was ran on the dataset. **greatest(See Figure 3)**

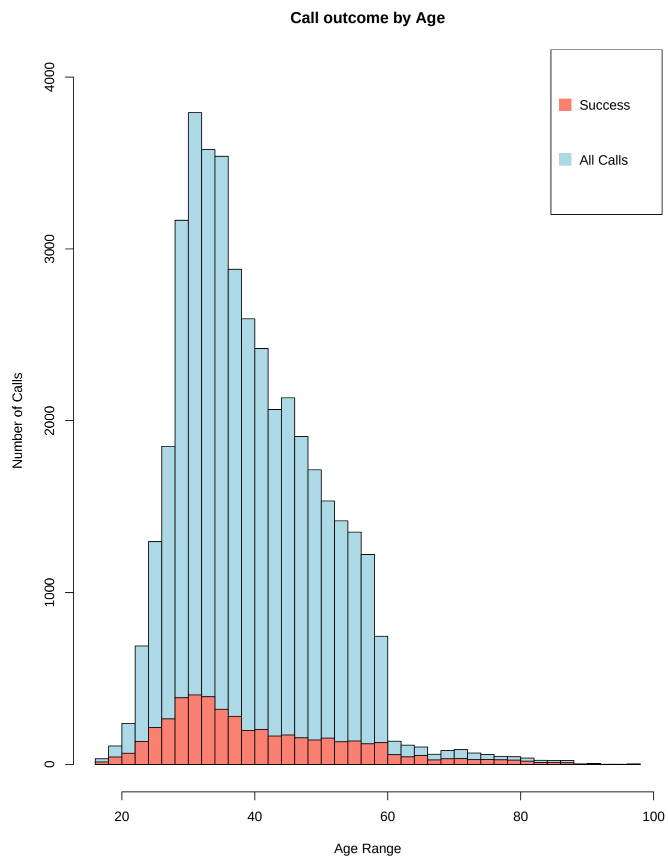


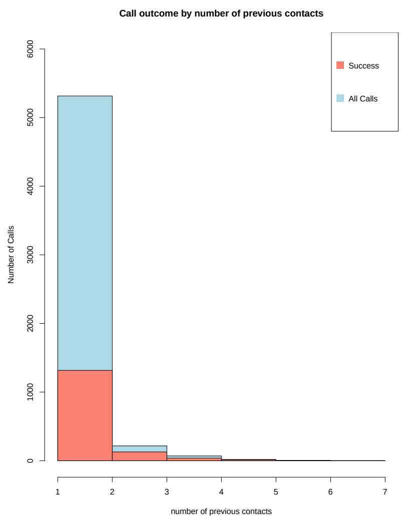
Fig. 4. Age

B. Data Exploration

After examining the quality of the data that we have and the individual correlation between the data and the outcome the author deduced the following from viewing the graphs. Little correlation in terms of success from the following variables Age **(see Figure 4)**, Job, Marital Status, Education, Credit Default, Housing Status, Loan Status,Contact Method (but there does seem to be a better outcome for cellular than telephone presumably landline) and Day of the week. However the following variables do show a correlation.

- Month - There is a clearly better outcome for calls made in periods when there were few calls than in those when there were many. I think that what is happening here is that customers who are engaged with the bank make subsequent calls and therefore these later calls appear to correlate with successful outcomes but this may be a false indicator as it may be data that is unavailable when selecting customers for calling.
- Previous Campaigns - There is a clear indication that a successful outcome in a previous campaign is a good indicator of a successful outcome here unfortunately there are very few positive previous outcomes.
- pdays - Number of days passed after last contact on a previous campaign 999 means no previous campaign.there is a clear indication that any previous contact is a good thing - but no particular correlation with the number of

days we may already have this data from the previous campaign indicator - so it may just be noise.
- Previous - There is already a clear indication that any previous contact is a good thing .Here we are seeing that higher number of previous contacts indicate better outcomes.**g (see Fig.5)**



Contacts.pdf

Fig. 5. Previous Contacts

V. METHODOLOGY

Machine learning algorithms work by searching through a set of possible prediction models for the model that best captures the relationship between the descriptive features and the target feature. Therefore, the goal is to seek models that are most consistent with the data and provide a good fit. Some models memorise the dataset, a process known as overfitting, which occurs in models that do not generalise beyond the dataset. It occurs when models are excessively complex. The converse occurs when a model is too simple, resulting in it generalising too much. This is known as under-fitting. The goal is to strike a balance between complexity and simplicity (over-fitting and under-fitting), which can be difficult to achieve. The author has two measure of success, number of calls converted to term deposits and number of calls made.The objective is to retain the maximum number of conversions while making the fewest number of calls. making very few calls at a good conversion rate but only picking up a few deposits is not particularly desirable. A better measure might be how many calls we can skip while keeping the missed deposits rate well below the unfiltered rate of 11.27%.

*

The current performance is as follows

yes : 4640 = 11.265% no : 36548 = 88.735%

*

Here we are gong to measure the performance of our classification based off of 2 benchmark models.

1) Random model - Which Sets the Call (Y/N) to Y for 80
2) Manual model based on the characteristic with the most obvious skew.This is better than random and is the real baseline we need to exceed to show we are doing something useful.

A. Data Analysis and Results

From examining the data we can get a pretty good result by excluding customers with an unknown credit status and a telephone (rather than cellular).This looses us 4233 calls but only 152 accounts (3.6% compared to 11.3%)Excluding all telephone customers looses 15044 calls but only 787 contracts (5.2%) This gives two useful ends to the results range. A lot of calls were lost (36%) but with only 17% of the accounts being lost in the process. Our model output is expressed in terms of a vector called Call which has the same number of elements as the original data (41188) All customers are set to 'yes' and then customers to be excldued are set to 'no' .**Figure 6** shows random selection of choosing 80% customer to call gives no benefit - i.e. we loose the same proportion of business as we make. Both bemchamark models produced a similar score with the Random model scoring 11.25% and the Manual model scoring 11.26%. Next models are based on observed imbalance between **credit status:unknown** and **call type:telephone** Conversion rate increases to 14.7% while while skipped calls would have converted to 787 deposits at at a rate of 5.2% which is relatively efficient. The same model used with telephone variable produced 12% and 3% respectively. So we have established two performance extremes based on observed data excluding a lot of customers with telephones gives many fewer calls but gives a ¿14% conversion on calls made excluding fewer customers based on telephone and unknown credit score gives a 12.1% conversion but looses very few potential deposits.

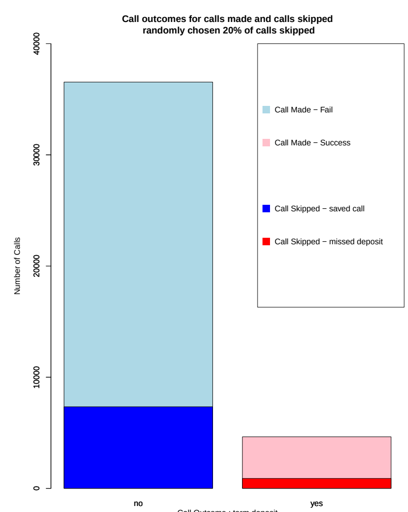


Fig. 6. Call outcomes for random 20%

B. Neural Networks Results

The dataset was divided into 25% training set and 75% test set. This is not too large for computive intensive models such as Neural Networks but large enough to avoid problems with sparse values and overfitting. The successful calls were deliberately given extra weight so the neural net will give them more consideration. After Neural Network prediction the author then derived a Call/Skip value from these results. The salient point from this is that the threshold we choose for skip is the key variable. The threshold chosen for this 0.5 which delivered a Call Success rate of 14.64%. The Neural Network works reasonably well given that the current performance is 11.64%. The key is to pick the threshold of acceptability and the output will give you a list of customers to call. The higher the threshold the fewer customers that will be called but the greater the conversion rate. Choosing the number is driven by the business value of the deposit account versus the business cost of making the calls. There is a significant bump in the output of the neural network as shown in **figure 7**.

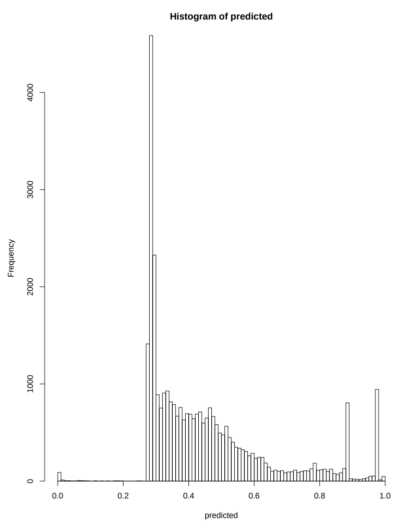


Fig. 7. Neural Network Prediction with Threshold of .5

A different value was used to try and eliminate this spike using an identical method as used previously. Weight was attached to the successful calls this time so the neural net will give them more consideration. The weights chosen reflect the yes/no balance in the data. After choosing a threshold of .40 a score of 17.54% was reached. This gives a smoother graph as can be seen in **Figure 8**. When the NN model was ran again differing results were noted with the threshold set to .4 giving a 23.41% "Call Success Rate" and a threshold of .5 giving a 18.45 success rate which is the inverse of the previous run. The results are pretty variable so the best thing would be to experiment with the weights until a favourable outcome is achieved. Also setting skip=True gives a much better result with far fewer poor results and no spikes in the histogram. In conclusion to achieve a better outcome using the Neural Network is to look at the range of values for the training data and re-run if necessary if there is a poor spread. Once a good spread of values has been obtained then just pick the target or budget and let the Neural Net give the customer list.

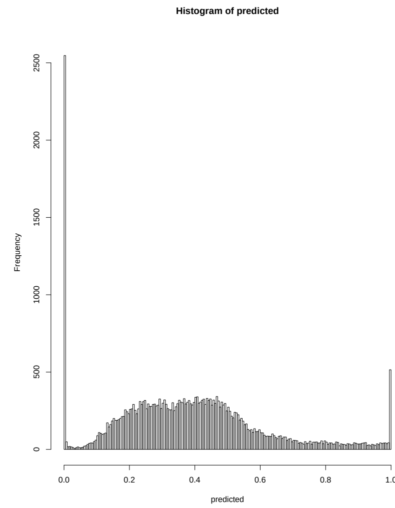


Fig. 8. Neural Network Prediction with Threshold of .4

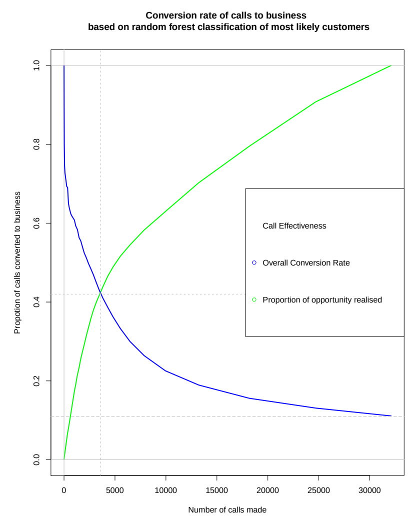


Fig. 9. Random Forest Conversion rate

C. *Random Forest Results*

The same process was carried out for the Random Forest model as for the Neural Network model except that a numerical representation of the outcome 0.0 (NO) - 1.0 (Yes) was used. This helps to train some models to give a value between the two limits which we can use as a way of selecting a larger pool of customers to call by choosing a low threshold. For this exercise its imperative to eliminate poor customers and not only call the very good prospects. As before the data was divided into a 25% training set. The Random Forest Predictor uses the same Call/Skip values as per the Neural Network prediction. As before the key variable is the threshold we chose for skip. The **"Call Success Rate"** using the Random Forest works reasonably well with a score of 22.55%. This is double the original "Call Success Rate" of 11.25%. As per the Neural Networks Predictor picking an appropriate threshold of acceptability is key. The higher the threshold the fewer customers called but the conversion rate so choosing the number is driven by the business value of the deposit account versus the business cost of making the calls.**(See Fig.9)** The Random Forest model identified most of the good customers early on **See Fig.10** but it misses almost all until the end using only 200 trees. When we use 2000 trees we get a "Call Success rate" of 23.22% **(See Figure 11)**

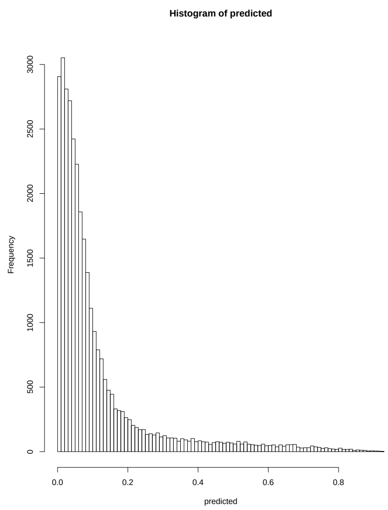


Fig. 10. Random Forest Conversion rate

D. *Random Forest Results*

Conclusion

The critical point of this exercise is to find a way to reduce the number of calls made but keep the number of successful deposit accounts created. Customers need to be classified into call/don't call groups. Producing a very small call list with an expectation of very good results is not the objective. The objective is to find a way of reducing the call list but keeping pretty much all the successful outcomes in the reduced list. There are two measures of success - what percentage of the successful calls did we keep - and how well did we eliminate unsuccessful calls while doing it. The difficulty in such a classification problem is that there is no metric in the data set to measure success as we don't want to only call customers with a high level of success but reduce the amount of calls made but only targeting customers with the greatest chance of success. With this in mind the performance of the algorithms will have a much different measure as there is nothing in the data set with the "right answer" for this problem. Therefore predicting with 90% plus accuracy is impossible so what we have to aim for is to predict how many calls we can skip while keeping the missed opportunities to a minimum using appropriate benchmarks. 11.25% is the unfiltered conversion rate. The success of the Neural Network model was varied with the greater prediction success coming when the threshold chosen was lowered ie. more customers called but at a lower conversion rate. The Random Forest is more reliable than the neural network which sometimes returns illogical results. However the random forest gives poor results towards the bottom end - so while it identifies most of the good customers early on it misses identifying them all until the end (i.e almost all calls made) - whereas the neural net is very good at identifying the poor conversion rates and leaves far fewer good customers in the lower priority calls.Overall recommendation with this type of data if planning a new telesales campaign based on the results of a previous one is Random forest is a very good predictor of where to get a good proportion of the available business for a small number of calls - whereas a neural net is better at identifying how to get a larger portion of the available business while minimising the number of calls required to achieve it. If this study where to be undertaken again with more time and resources a Support Vector machine would be explored also.

Acknowledgment

National College of Ireland Urvesh Bhowan Noel Cosgrave

References

References

[1] Moro, S., Laureano, R. and Cortez, P., 2011. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In Proceedings of European Simulation and Modelling Conference-ESM'2011 (pp. 117-121). EUROSIS-ETI.
[2] Moro, S., Cortez, P. and Rita, P., 2014. A data-driven approach to predict the success of bank telemarketing. Decision Support Systems, 62, pp.22-31.
[3] Ngai, E.W., Xiu, L. and Chau, D.C., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. Expert systems with applications, 36(2), pp.2592-2602.
[4] Bahari, T.F. and Elayidom, M.S., 2015. An efficient CRM-data mining framework for the prediction of customer behaviour. Procedia computer science, 46, pp.725-731.
[5] Olson, D.L. and Chae, B.K., 2012. Direct marketing decision support through predictive customer response modeling. Decision Support Systems, 54(1), pp.443-451.
[6] Thearling, K., 2017. An introduction to data mining.
[7] Martens, D., Vanthienen, J., Verbeke, W. and Baesens, B., 2011. Performance of classification models from a user perspective. Decision Support Systems, 51(4), pp.782-793.
[8] J. Moeyersoms, D. Martens, Including high-cardinality attributes in predictive models: a case study in churn prediction in the energy sector, Decis. Support. Syst. 72 (2015) 7281.
[9] Asif, M., 2018. Predicting the Success of Bank Telemarketing using various Classification Algorithms.
[10] Beucher, A., Mller, A.B. and Greve, M.H., 2017. Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark.
[11] Coussement, K., Lessmann, S. and Verstraeten, G., 2017. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. Decision Support Systems, 95, pp.27-36.
[12] Elsalamony, H.A., 2014. Bank direct marketing analysis of data mining techniques. International Journal of Computer Applications, 85(7).

VI. Appendix

	age	job	marital	education	default	housing	loan	contact	month	day_of_week	pdays	previous	poutcome	y	ynum
1	56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	999	0	nonexistent	no	0
2	57	services	married	high.school	unknown	no	no	telephone	may	mon	999	0	nonexistent	no	0
3	37	services	married	high.school	no	yes	no	telephone	may	mon	999	0	nonexistent	no	0
4	40	admin.	married	basic.6y	no	no	no	telephone	may	mon	999	0	nonexistent	no	0
5	56	services	married	high.school	no	no	yes	telephone	may	mon	999	0	nonexistent	no	0
6	45	services	married	basic.9y	unknown	no	no	telephone	may	mon	999	0	nonexistent	no	0
7	59	admin.	married	professional.course	no	no	no	telephone	may	mon	999	0	nonexistent	no	0
8	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	999	0	nonexistent	no	0
9	24	technician	single	professional.course	no	yes	no	telephone	may	mon	999	0	nonexistent	no	0
10	25	services	single	high.school	no	yes	no	telephone	may	mon	999	0	nonexistent	no	0

Fig. 11. Full Dataset

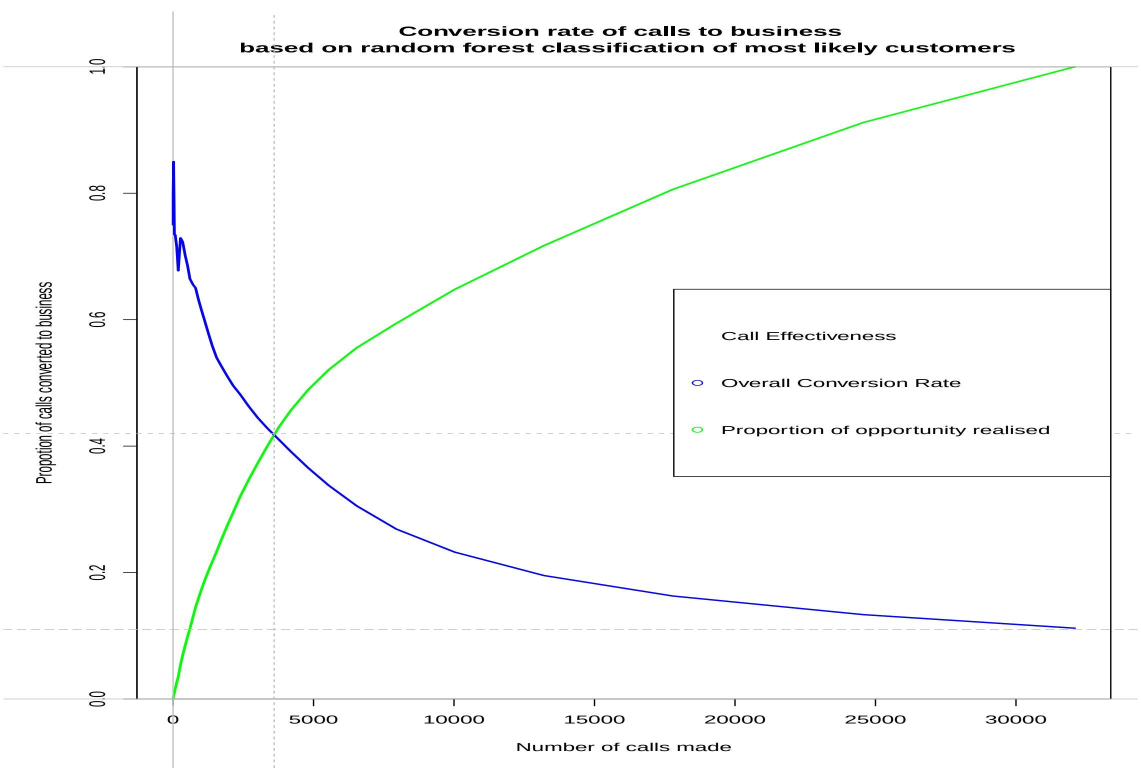


Fig. 12. Conversion rate using 2000 Trees

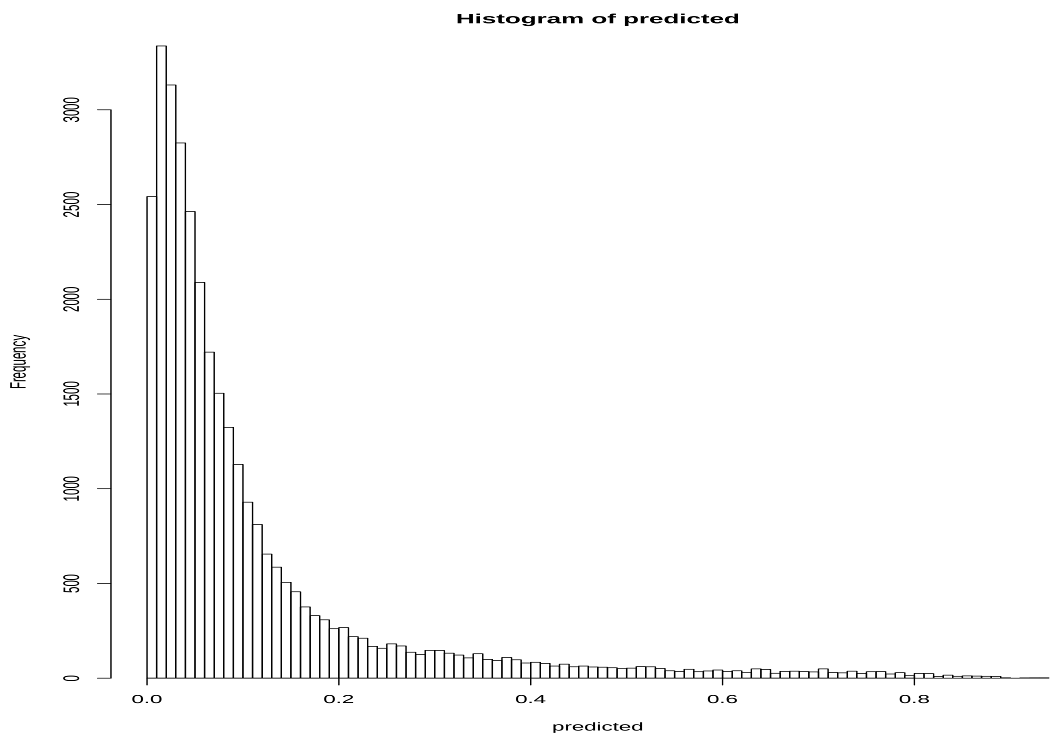


Fig. 13. Random Forest using 2000 Trees