Predictive Analytics in Marketing

A Practical Example from Retail Banking

by Alvin Choong,

with input from David Menezes, Frank Devlin, Mudit Gupta, Tan Wei-Chyin and Kate Chen

ABSTRACT

In this research note, we present a case study where we have applied predictive analytics methods to a historical retail banking dataset in order to yield usable insights for marketing of a term deposit investment product. The primary focus of the paper is to illustrate the typical framework and concepts used by data scientists within the machine learning paradigm. In doing so, we hope to demystify the approach to fledgling actuaries approaching predictive analytics for the first time. Links to the source code used to create the results in this paper is provided.

KEYWORDS: predictive analytics, machine learning, generalised linear models, gradient boosting models, training, validation, R.

1 Introduction & Motivation

The technical skills of an actuarial practitioner can be of considerable value in a wide range of areas, beyond insurance, pensions, and investment – i.e. areas traditionally associated with actuarial science.

One area where actuarial involvement is gaining ground is Big Data, and an increasing number of actuaries are utilising their statistical and programming skills to make strides in the predictive analytics and machine learning fields. In fact, these fields are not new to actuaries. A sizeable proportion of general insurance actuaries and a number of life insurance actuaries, through personal lines pricing and mortality analyses, have been using generalised linear modelling techniques in their respective areas for a number of years now.

As part of our research remit, the Singapore Actuarial Society (SAS) Big Data Committee is pleased to present this case study, which is the first in a series of articles. This note illustrates how predictive analytics can be applied to a historical banking dataset in order to yield usable insights for marketing. The ideas presented in this case study can be applied in other contexts outside of banking; for example, in insurance one must answer questions such as, what is a suitable product to recommend to a customer? What is the best time to market this product? And, which is the most effective channel to contact a customer? In preparing this study we have sought to provide clarity on how such techniques may be applied while also addressing some of the real-world issues that you may encounter within a commercial context.

2 Data for the Case Study

We have based our case study on the dataset used by Moro et al., 2014[1]. This database details the results of a telemarketing campaign held by a Portuguese banking institution to sell a specific term-deposit product. It has been made publicly available and may be downloaded from the following URL: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Readers of this white paper that may have been participants in the SAS's Asia Actuarial Analytics Challenge 2017 will notice similarities between the dataset above and the competition's dataset. This is no coincidence – the dataset above was been sampled and altered by the Committee for the purpose of this year's Kaggle competition.

To view the competition, please visit the following link: https://inclass.kaggle.com/c/asia-actuarial-analytics-challenge-2017. The competition ended on 30 September 2017, but it is still possible to participate in the forums and compete against the leading models.

Some details of the dataset are provided below:

- The dataset contained 41,188 individual records. Each record represents a call that was made to a bank customer.

[1] [Moro et al., 2014] 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

- For each record, the call operator recorded if a particular term deposit[2] product was subscribed. In technical jargon, this may be considered the *response variable* (or *y-variable* in regression statistics).

- In addition, for each record 20 features (or *x-variables* in regression statistics) were captured. These features include customer-specific variables such as the age of customer; the customer's occupation; marital status; education level; and when the customer was last contacted by the bank prior to the call. External variables were also provided such as the weekday in which the call was made; the month in which the call was made; and even the prevailing interest rates at the time the call was made.

- The dataset is ordered chronologically and covered the period from May 2008 through to November 2010.

3 Objective of the Case Study

This case study has two goals. The first is to show how a simple predictive analytics exercise may be carried out on real data using open source statistical modelling software. The second goal is to demonstrate how the results from a predictive analytics study can be applied to produce real, tangible improvements in a company's business performance. This case study therefore places less emphasis on model accuracy in a pure statistical sense, and instead focuses on the practical aspect of using the results effectively, addressing the *"so what?"* question that often accompanies such results.

In this case study, we aim to predict, given relevant customer information, whether a particular customer will choose to subscribe to a term deposit product. But how would such a prediction be of use to the bank in question?

In general, direct marketing campaigns tend to require dedicated in-house or out-sourced call centres. The cost of running such sales teams can put a considerable strain on the expense ratio of the product. A targeted approach in which the company is able to identify customers who are more likely to subscribe is desirable and would allow greater focus on those customers most likely to generate a sale. In addition to being more efficient, the potential reduction in marketing costs is likely to increase the profit margin of the product overall. While this case study specifically considers sales in the context of call centre teams, the principles outlined are generalised and could equally be applied when trying to improve conversion ratios for online or mailshot sales channels.

Consider the following simple example. Suppose we need to contact 200,000 customers. Each call made to a customer costs $5 and we have a total of $1m to spend on marketing. Let us assume that the bank generates a profit of $100 for each successful product sale (i.e. without allowing for marketing costs), and from historical experience we know that 1 in every 10 calls results in a successful sale.

[2] A term deposit is a deposit held at a financial institution that has a fixed term. These are generally short-term with maturities ranging anywhere from a month to a few years. Interest is typically paid at maturity.

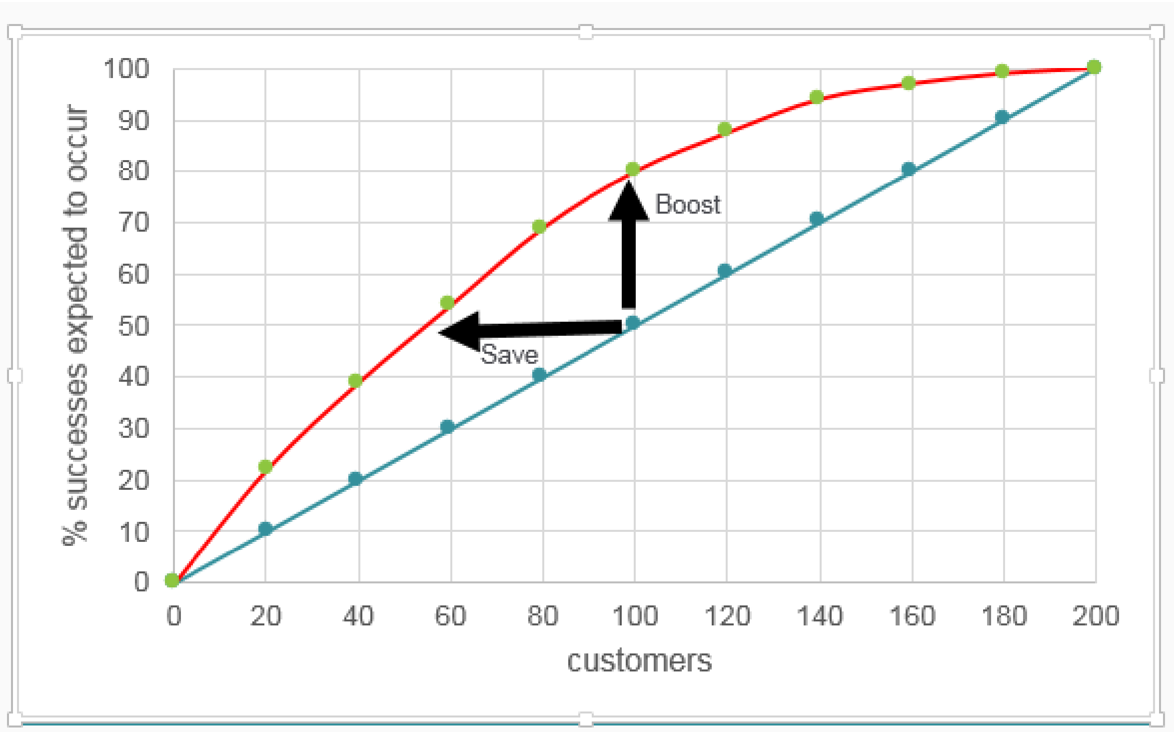
If we were to call 100,000 of our customers randomly, we would expect to have 10,000 sales, and $0.5m in profit after allowing for telemarketing costs ($1m minus $0.5m). This is represented by the blue line in Figure 1 below. But can we do better than this?

With machine learning, the answer is a resounding yes. On average customers are 10% likely to respond favourably to a marketing call, but in reality this probability varies from one customer to the next. If we could somehow find a way to rank customers from "most likely to buy" (say 50%) to "least likely to buy" (<5%), we could easily increase savings or profits by focusing on the right customers. This is where our predictive models give us tremendous added value.

If our model succeeds in differentiating customers effectively (see the red line in Figure 1 below), we can obtain 80% of the successful sales from making the same number of calls as before (100,000). This would give us 16,000 (20,000 x 80%) successful sales, and a profit of $1.1m ($1.6m minus $0.5m) after telemarketing cost. In this case, we have boosted our profits by more than double (2.2x) with the help of this predictive model.

Alternatively, if our emphasis is to make fewer calls but achieve the same volume of sales, by using the predictive model we could achieve 10,000 product sales by calling just 55,000 customers. This would mean a marketing expenditure of only $0.28m, and we will have saved on marketing expenses by almost a half.

Figure 1: Graph showing how the cumulative % of successes varies with the number of customer calls made. The red line represents a predictive model, the blue line represents random, i.e. where no predictive model has been used.



Therefore an accurate predictive model, if applied correctly, gives the bank the potential to improve financial performance significantly.

Apart from the financial significance illustrated above, depending on the actual methodologies used, we can also infer and draw useful insights on the relationships between the features and the response variable. The inferred relationship could then be utilised in other commercial decision-making processes of the company. In this particular example, modelling results may suggest that older customers are less likely to subscribe to a term deposit than younger customers. Knowledge of these types of relationships could be applied at the product design stage and new products which appeal more to older customers may be developed.

4 Modelling Approach

4.1 *Software*

Our case study was carried out using the R programming language[3], together with the *caret*[4] package (*caret* is short for classification and regression training). R was selected on the basis of:

- Its increasing popularity with universities running statistical and actuarial courses;
- The availability of pre-compiled packages, and
- The fact that it is free.

Similar models can be built using other widely available programming languages such as Python, SAS, MatLab or Octave.

The *caret* package was selected for the case study as it contains numerous ready-made functions that enable a large number of complex regression and classifications models to be run using the same few lines of code. This makes it easy to follow and therefore suitable for an introductory case study. Within *caret*, we selected several commonly used models for the purpose of illustration. It is worth noting that these models were chosen for the purpose of this study and not for any specific reasons related to their statistical properties.

4.2 *Data processing*

For the purpose of recreating a real-world situation, we first put aside 20% of the most recent information, and assume that the outcomes of these are not known at the time of modelling, i.e. they occur in the future at the time of modelling.

In machine learning, this dataset is known as the *test* dataset, and is the dataset we are required to predict. Then, consistent with generally-accepted machine learning practices, we segmented the remainder of the calls into two separate datasets – a *training* dataset and a *validation* dataset using a 75:25 ratio. The split between *training* and *validation* datasets is done randomly.

The training set was used to fit a number of alternative predictive models. These models were designed to predict the probability that customers called in the future (i.e. from the *test* dataset) would subscribe to the term deposit product. Having built our series of models, the validation dataset was then used to estimate the prediction error associated with each model. This prediction error measure would form the basis for selecting the preferred model.

The *validation* dataset is created in order to prevent overfitting[5] from occurring during the modelling process. Without the *validation* dataset, there is a risk of fitting a model that has too many

[3] https://cran.r-project.org/
[4] https://cran.r-project.org/package=caret
[5] https://en.wikipedia.org/wiki/Overfitting

parameters or is overly complex. Although such a model would fit extremely well to the *training* dataset, it would tend to allow for "noise" or random error that is specific to the *training* dataset alone, leading to poor predictive performance whenever this model is used on new *test* data.

Finally, based on the response data collated in respect of the *test* data we were able to determine the accuracy of the final chosen model against future data. A summary of these splits is shown below in Table 1:

Table 1: Data Splitting

Dataset	No of records	Proportion of Records
Full	41,188	100%
Training	24,713	60%
Validation	8,238	20%
Test	8,238	20%

One key observation to note about the dataset is that only 11% of all calls result in a sale (i.e. classified as "yes"), and the remaining 89% do not (i.e. "no"). This is an example of an *imbalanced data*, as the classes here ("yes", "no") are not represented equally in the dataset. The imbalanced nature of the dataset may result in model accuracy being incorrectly measured, depending on the metric used. As an example, a model could be correct 9 times out of 10 simply by predicting "no" on every test record, but such a model would provide no practical value.

This case study does not include discussions relating to the handling of imbalanced data, although this issue was considered in the selection of the model evaluation criteria, further details on which is discussed in Section 4.5 below. In practice, there are a number of ways one may allow for imbalanced datasets within the modelling process[6].

4.3 Exploratory Data Analysis and Selection of Features

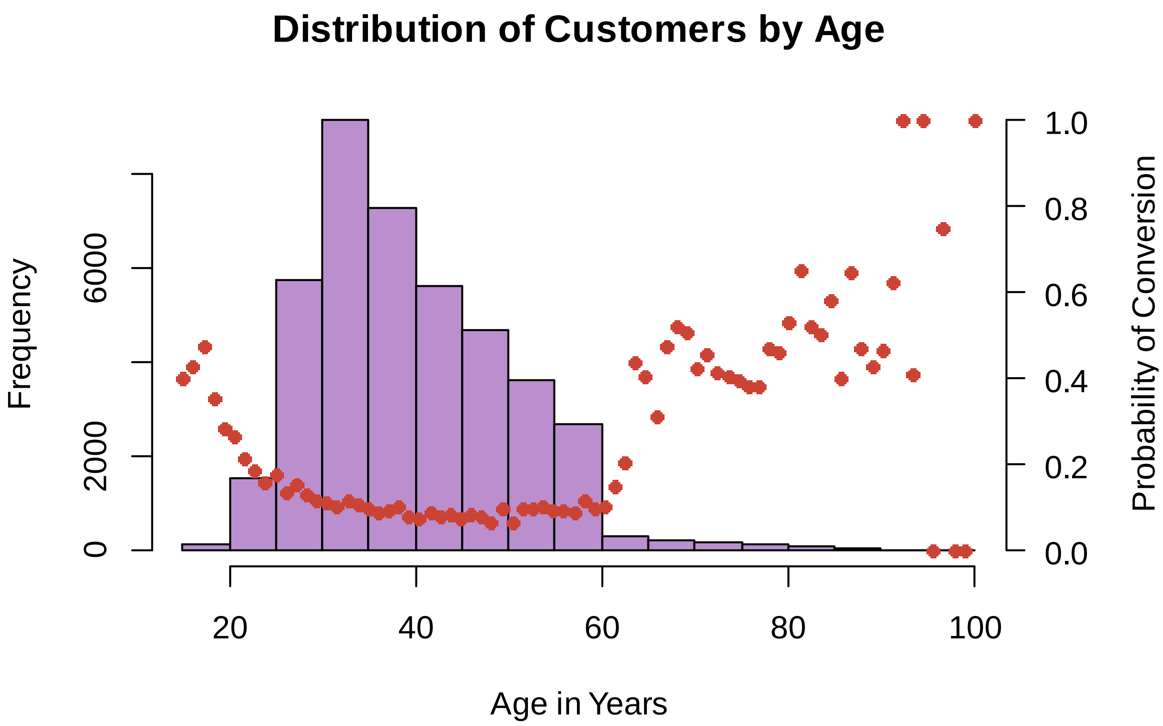
The term *"feature engineering"* may be defined as the process of using domain knowledge of the data to create features that make machine learning algorithms work[7]. In the context of this case study, we carry out feature engineering by applying transformations to the variables in the dataset. Transformations typically involve grouping a large number of levels into a smaller number of levels (e.g. grouping ages in individual years into bands of 10 years), or creating new features out of existing features (e.g. creating a day of week variable from a date field).

The feature engineering carried out in this case study was based on a series of exploratory analysis, consisting mainly of one-way summary tables and visualisation graphs. Modifications which were particularly useful include the grouping of continuous variables such as interest rates and consumer price indices into several buckets, which ensures a more credible volume of data in each variable level. Histograms and barplots were also useful for this purpose.

[6] https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/
[7] https://en.wikipedia.org/wiki/Feature_engineering

Three examples of such plots are set out below in Figures 2, 3 and 4 below:

Figure 2: Illustration of age vs conversion rate



Here we see that the probability of conversion decreases with age up to the age of 40. Between 40 and 60, the probability of conversion remains stable, with no clear trend either way. Beyond the age of 60, the sample size reduces significantly, resulting in a high degree of volatility or "noise". In this case, a decision may be taken to carry out feature engineering, e.g. group the age factor into a smaller number of age groups in order to increase the statistical credibility of the modelling result. For instance, one may select to group all ages 60 and above into a single age group. Note that we are not making a causal connection between age and probability conversion – rather the inference is that, all things being equal, up to the age of 40 older customers were on average less likely to purchase the term deposit product compared to younger customers.

Figure 3: Illustration of customer job vs conversion rate

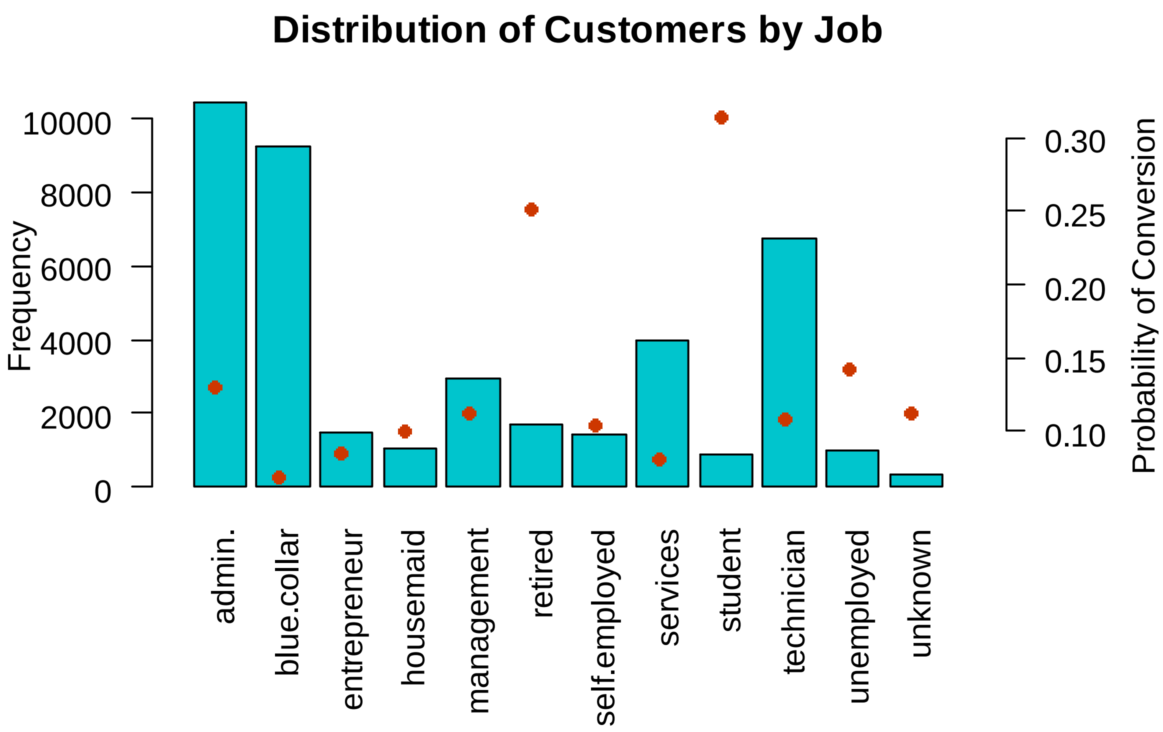
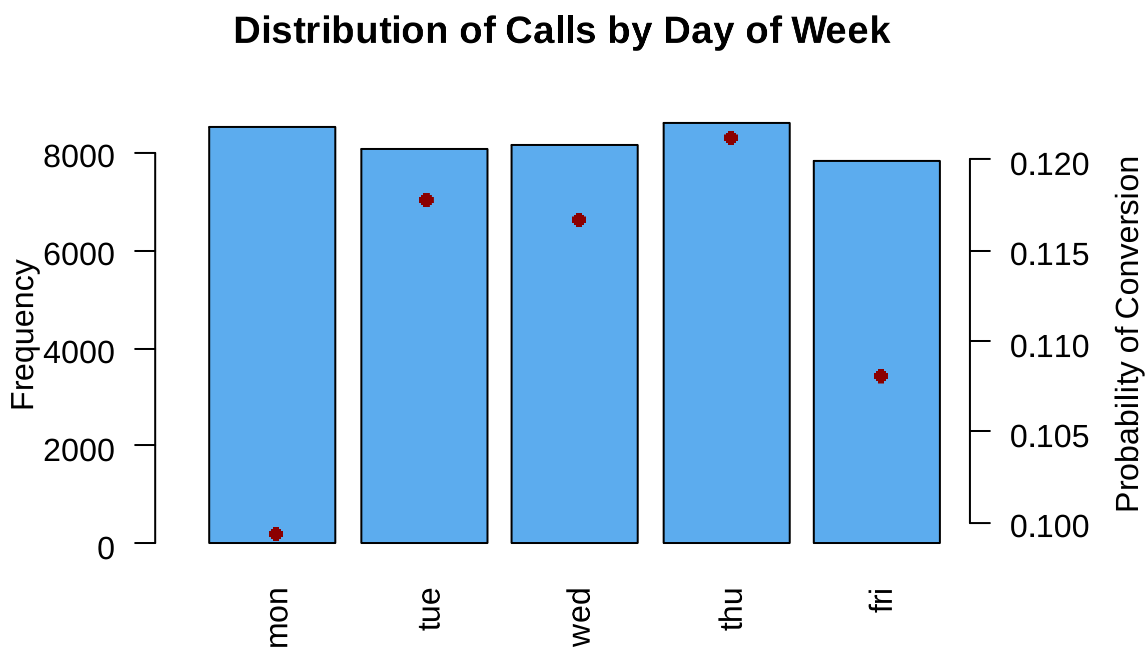


Figure 3 shows how the probability of conversion varies with the customer's occupation. Students and retired customers were most likely (>25%) to purchase the term deposit products, while blue collared workers and entrepreneurs were considered least likely (<10%) on average.

Figure 4: Illustration of weekday of call vs conversion rate



Similarly, Figure 4 shows how the conversion probability varied by the weekday during which the call was placed. Raw data suggests that, all things being equal, customers were more likely to respond favourably when called in the middle of the week, compared to Mondays and Fridays. The difference in conversion probability between the best day (Thursday) and the worst day (Monday) was higher than 2%, which is material in the context of product sales.

Through these simple exploratory graphs, a number of features such as the three discussed above were identified to be useful features in that including them in a model were likely to result in the model being more predictive. Some other features were found to be less predictive than the ones shown above and modellers may ultimately decide to exclude them from their final models.

It is critical to note that the exploratory graphs above are on a "one-way" basis, in that they do not consider the impact of correlation with other features, and are therefore only meant to provide indications on how useful the feature might be in determining conversion probabilities. A feature found to be useful through exploratory analysis may subsequently lose its predictive importance when combined with other highly correlated features within a model. More reliable measures of importance are those that are based on model outputs, as shown in Section 7 below. The multicollinearity of attributes should be tested with a scatterplot matrix[8], as correlation of variables can be problematic for some techniques (e.g. GLM).

It is also important to note that this data contains a potential modelling pitfall which arises from one of its features. The *duration* feature refers to the duration of the last contact with the customer.

[8] https://wiki.smu.edu.sg/anly104/ANLY104_T8_Data

However, as pointed out in the dataset description, the duration is only known after a call has been carried out, and in all cases where duration equals zero, y is also zero (i.e. there has been no sale to that customer). The duration factor would therefore not be useable in reality, and would have no practical application in building an effective predictive model.

However, if we had overlooked this issue and instead selected to include it within our predictive models, we would end up with a very different modelling outcome. Due to how this information was recorded (duration is zero if y=0, and non-zero if y=1), the duration factor would be extremely predictive of the final outcome, and any model that included this factor would very accurately predict the outcome of each call. This would give us an extremely accurate model as measured by any model evaluation criteria, but ultimately one with little practical value as it cannot be used in reality.

In Section 5 below we show how other studies that include the duration factor show a level of accuracy that is much higher than would have been obtained otherwise. In many respects, the risks of making this error is exacerbated by its effect of overstating model accuracy – a high model evaluation score may lead a modeller to erroneously conclude that his or her model was appropriately constructed.

This emphasises the importance of adopting a holistic approach when performing a predictive modelling study – one may not obtain useful results simply by applying modelling techniques without first understanding the underlying data. Domain expertise also helps to increase the level of understanding around the modelling data.

4.4 Model Fitting

Following the feature engineering and selection processes, several models from the *caret* package were fitted to the *training* dataset, and tested against the *validation* dataset. In our case study, we have only selected a small number of models for illustrative purposes.

Using a chosen selection criteria (further elaborated below), we selected the best-fitting model and fitted this to our *test* dataset, which related to the future time period (most recent 20% of calls). Finally, we evaluated the accuracy of our predictions.

4.5 Model Selection

We fitted a range of models from the *caret* package to the training dataset, namely:

1. GLM (Generalised Linear Models) – the logistic regression was used in this case as the response variable is binary (0/1, or No/Yes).[9]
2. Boosted GLM [10]
3. Regression trees, also known as Classification and Regression Trees (CART) or decision trees [11]
4. Stochastic Gradient Boosting (GBM) [12]
5. Random Forest [13]

[9] More information on logistic regressions: https://en.wikipedia.org/wiki/Logistic_regression
[10] More information on boosting: https://en.wikipedia.org/wiki/Boosting_(machine_learning)
[11] More information on trees: https://en.wikipedia.org/wiki/Decision_tree_learning
[12] More information can be found at: https://en.wikipedia.org/wiki/Gradient_boosting
[13] Introduction to random forests: https://en.wikipedia.org/wiki/Random_forest

6. Conditional Inference Trees [14]
7. Multivariate Adaptive Regression Splines (MARS)[15].

Details on the models above are provided in the footnotes below. Further details may be found from external sources online.

In addition to the 7 models above, a further two *ensemble* models were derived from a combination of the 7 models above. An ensemble model is constructed by combining multiple related but diverse models, in order to generate a 'blended' model that usually achieves a greater level of predictive accuracy than the original models.

For illustrative purposes, we have performed basic forms of ensembling:

- Ensemble 1, consisting of a simple average of the following 4 randomly chosen models – GBM, random forest, conditional inference trees, and logistic regression models.
- Ensemble 2, consisting of a weighted average of the same 4 models above, but with higher weighting being assigned to better-performing models.

For this case study, we only included 7 models and two (simple) ensemble models, and only used one evaluation criteria to select the best-fitting model. This simple approach to modelling was chosen to increase the clarity of this study, and to demonstrate how a predictive model may be built from a given dataset. In practice, we would tend to fit a wider range of predictive models with a greater focus on feature engineering, and our model selection process would typically involve a higher level of sophistication.

In order to assess the accuracy of each model, we evaluated each model against the *validation* dataset using the Area under Curve (AUC) metric (the evaluation is done on the *validation* dataset to prevent overfitting, as described in Section 4.2 above).

The AUC metric is commonly used to evaluate models with binary outcomes, and is a useful metric to use with an *imbalanced* dataset such as this case study. A purely random predictive model (i.e. coin toss), would give an AUC figure of 0.5, and a perfectly predictive model would give a figure of 1. The closer the AUC is to 1, the more effective the model is deemed to be in terms of predicting future data. A good detailed explanation on the AUC metric can be found in the following link on Stack Exchange's Cross Validated website: https://stats.stackexchange.com/questions/132777/. A cost matrix to allow for different costs of different outcomes can also be included.

5 Modelling Results

Table 2 below sets out the AUC values of the 9 models above, which have been trained on the *training* data. These AUC values measure the fit of these trained models in respect of the *validation* data, i.e. how well each trained model predicts whether or not a customer in the *validation* data subscribes to the term deposit product.

[14] Difference between CIT and traditional trees: https://stats.stackexchange.com/questions/12140/
[15] More info on MARS model: https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_splines

The table below is ordered in descending order of AUC, so the higher the position of the trained model, the better its performance relative to the *validation* data.

Table 2: Model fit results

Model Sequence	Model	AUC
9	Ensemble 2 (Weighted Average of 1,4,5,6)	0.656
8	Ensemble 1 (Simple Average of 1,4,5,6)	0.653
7	Multivariate Adaptive Regression Splines	0.651
3	Regression Trees	0.650
6	Conditional Inference Trees	0.646
4	GBM - Gradient Boosted Trees	0.646
2	Boosted GLM - Logistic Regression	0.638
1	GLM - Logistic Regression	0.633
5	Random Forest	0.608

The second ensemble model ("Ensemble 2") appears to give the highest AUC value when applied to the *validation* dataset. If we were to select a model based purely on statistical fit, we would have picked this as the model of choice. In practice however, other considerations such as model complexity, communicability and scalability would come into play, and simpler models such as the logistic regression or GBM models could be selected instead.

It is worth noting that, similar to many traditional areas of actuarial work, there is process uncertainty inherent in the results set out above. Randomisation has been used at various stages of the modelling process, for example in the creation of training and validation datasets, and some algorithms such as the random forest (as the name suggests!) uses random numbers within the model-fitting process itself. In order to have results that are entirely reproducible, it is common practice to set fixed random seeds at relevant stages of the modelling process.

Although AUC values will tend to vary with different seeds, the variation is not usually significant such as to change the main conclusions drawn from the study. As a form of model validation, it may be useful to select alternative seeds to test for stability in modelling results. We have considered the sensitivity of modelling outputs to different seed numbers, but have not included the quantification of uncertainty within the scope of this case study.

Revisiting our point on the duration factor in Section 4.3, whilst a high AUC measure is desirable, it is also important to validate or investigate models with an exceptionally high AUC value. As part of this case study, the Committee conducted a research into various publicly available academic studies performed on the same dataset. In several of these studies, such as Chen et al. (2014)[16], Vaidehi (2016)[17], Alhakbani and al-Rifaie (2016)[18], and Elsalamony (2014)[19], the authors returned accuracy values that were significantly higher than the case study (above 0.90 in some cases). However, these

[16] J Chen, Y Han, Z Hu, Y Lu, and M Sun: Who Will Subscribe a Term Deposit? December 2014 (http://www.columbia.edu/~jc4133/ADA-Project.pdf)
[17] R Vaidehi: Predictive Modelling to Improve Success Rate of Bank Direct Marketing Campaign, March 2016 (http://www.ijmbs.com/Vol6/1/4-vaidehi-r.pdf)
[18] H A Alhakbani and M M al-Rifaie: Handling Class Imbalance in Direct Marketing Dataset using a Hybrid Data and Algorithmic Level Solutions, July 2016 (https://research.gold.ac.uk/17248/1/2016_SAI_Computing_IEEE_Class_imbalance.pdf)
[19] H A Elsalamony: Bank Direct Marketing Analysis of Data Mining Techniques, January 2014 (http://research.ijcaonline.org/volume85/number7/pxc3893218.pdf)

results arose from models which included the *duration* factor as a model feature. Since the duration field is populated based on the outcome of calls that have already been made, its inclusion in the model would tend to boost any measure of model accuracy significantly. Inclusion of the duration factor within the authors' models may have been deliberate and in line with the authors' objectives (for example demonstrating use of machine learning techniques to analyse historical information), but as our case study aims to predict outcomes based on information known prior to a telephone call occurring, it is imperative that the duration factor is not included so as to ensure the model results can be applied in a real-world setting.

This emphasises the value of understanding the data being modelled at the outset, ensuring the modelling approach is consistent with the intended use of the model, and also the importance of model validation and testing to ensure that the appropriate factors have been included.

6 Turning Results into Dollars and Cents

So how can we use our modelling results above to reduce telemarketing costs?

By applying our selected model(s) to our future *test* dataset, we can derive a score for each customer, which represents the likelihood of successful sale. We can then focus our telemarketing efforts on customers with high scores, and therefore reduce focus on customers who are deemed to have a relatively low probability of subscription. In this case, running the "Ensemble 2" model on the *test* dataset yields an AUC score of **0.646.** This is similar to the AUC obtained on the *validation* dataset above.

We can estimate the extent of profit increases as follows.

Table 3: Savings achieved using Ensemble 2 model on future test data

Number of future test records	8,238
Number of successful sales / conversions	2,540
Conversion rate	**31%**
50% customers called randomly	4,119
Number of successful sales / conversions	1,270
Conversion rate	**31%**
50% customers called based on Ensemble 2 model	4,119
Number of successful sales / conversions	1.580
Conversion rate	**38%**

A count of converted calls in the future *test* data indicates a conversion success rate of 31%, and that if we were to randomly call 50% of our *test* customers, we would on average obtain 1,270 sales (2,540 x 50%). However, by applying the selected "Ensemble 2" model on the *test* dataset, and then calling customers in descending order of expected conversion probability, we would have sold 1,580 policies, i.e. 25% more sales.

An alternative way to present this is to say that each call placed is 25% more likely to generate a sale using the selected model, which in the context of telemarketing is a significant improvement in success rate. It is worth noting that this significant improvement was achieved using a fairly simple predictive model with minimal feature engineering or model-tuning – the benefits would be even more pronounced if a more accurate predictive model (AUC of 0.80 and above) had been used instead.

The next step from here is to collect further details to help us in our quantification of profitability. Using additional information such as premium and profitability of the product as well as the cost per call, we can carry out the same analysis we set out in Section 3 above to accurately quantify the value of the predictive model in dollars and cents.

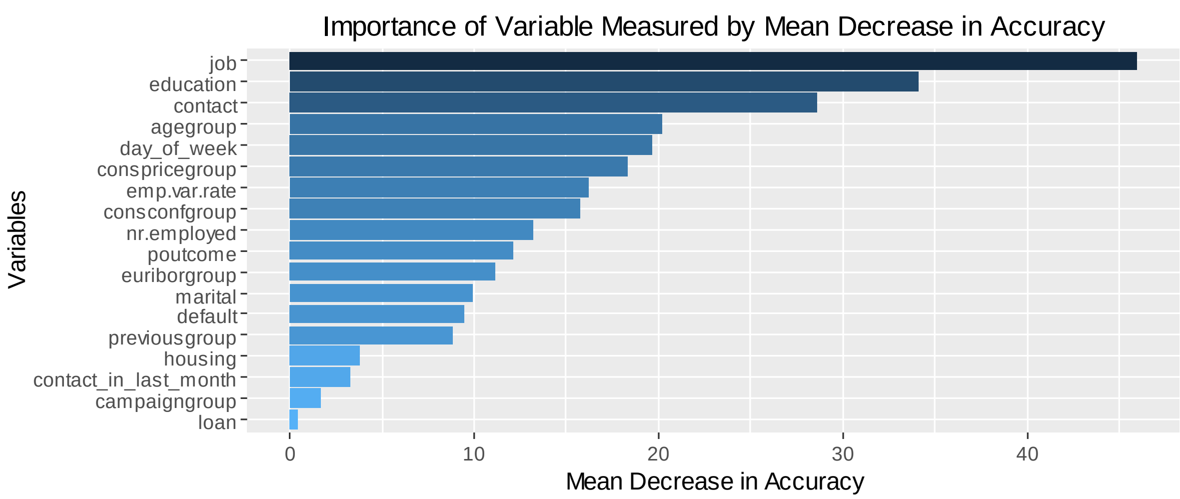
In practice, it would also be useful to carry out additional modelling exercises to further enrich the results of the study. For example, clustering analyses may indicate that there are distinct segments of customers – perhaps split by age or income – which may necessitate a different modelling approach (or even product offering) altogether.

We must also not discount the value that domain expertise, in this case term product experts and advisors, can add to the modelling process, for example to explain why specific features of the data behave the way they do, or even validate the results of the model through their own experience and knowledge of the bank's various customer profiles.

7 Which are the Most Important Factors?

Using the output of the fitted models, we can identify the most important variables in the model in terms of predictive power. One possible tool for this purpose is the *Variable Importance Plot*. As an illustration, we include a variable importance plot in Figure 5 based on the fitted random forest model:

Figure 5: Variable importance plot



Variable importance plots show how the model's accuracy reduces as each variable is removed from the model. For example, based on Figure 5, we can see that the customer's occupation and education are most important in terms of the random forest model's accuracy. This is consistent with what may be expected, given that occupation and educational level are likely to correlate strongly with the customer's bank balance, and therefore the level of disposable income available for investment in a term deposit product. We recall from Figure 3 above that our initial exploratory analysis had indicated that job may be a useful feature in terms of predicting conversion probability.

Perhaps a less expected outcome is the high predictive power of the day of week variable, which similar to occupation, had been identified as a useful model feature during the exploratory analysis. The models here suggest that calls made on particular days of the week, notably Monday, are less likely to result in a successful sale, all other things being equal. This is an example of an insight that may not have been particularly obvious without the help of our models. Such a result may suggest a need for further investigation to establish whether the observed impact of the variable is genuine, and if so, reasons for this should be identified. Similarly, these types of insights would enable the bank to take various steps to improve efficiency, for example by encouraging part-time workers to take Mondays off!

8 Conclusions and Final Points

Although the scope of the study above is constrained by the volume and availability of data – for example, the data does not provide an indication of the price and profitability of the product being marketed or the cost of making each call – it clearly demonstrates that predictive analytics can contribute towards increasing efficiency; reducing the marketing cost of the product and improving profitability as a result. More generally, it can be used widely and effectively to solve a large number of real-world commercial problems.

A final point to make is that, like all actuarial models, a predictive modelling exercise should not be thought of as a one-off exercise as relationships between variables and customer behaviour may change over time. It is imperative, once a model has been selected, to continually update the model over time with emerging experience, to continue enhancing the model by adopting other aspects not previously considered, and to incorporate domain expertise where possible.

This is true even in the specific case of our study above. The historical data used in the Case Study coincides with the 2008 financial crisis, which could mean that the results are biased to some degree by the prevailing interest rate environment that may be unrepresentative of what we would observe in a ‘normal’ scenario. A possible point for development therefore, could be to consider time-series effects with the modelling data itself, and to continue fine-tuning the model with new data over coming years.

Singapore Actuarial Society Big Data Committee

The objectives of Singapore Actuarial Society is to uphold the highest professional standards among members and to promote the study, discussion, publication and research into the application of economic, financial and statistical principles to practical problems in multi-disciplines with particular reference to Singapore and ASEAN region.

The Big Data Committee is the Society's initiative to explore the future of big data, analytics and unstructured data in Asia and what actuaries need to do to have the right skillsets that will be in demand for such work. The committee is made up of actuaries and data scientists based across Asia and globally from diverse range of industries. For enquiries please get in touch with Mudit Gupta (Mudit.Gupta@sompo.com.sg).

Appendix – Replication

- The following version of R was used to produce these results: R version 3.4.0
- The following version of the R GUI, R Studio, was used: Version 1.0.143
- The source code used in R can be found here: https://github.com/SASBDC/Bank-marketing/
- The following R packages were used:

    - **caret**
      Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, All an Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and T yler Hunt. (2017). caret: Classification and Regression Training. R package version 6.0-77. https://CRAN.R-project.org/package=caret

    - **randomForest**
      A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

    - **pROC**
      Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77. DOI: 10.1186/1471-2105-12-77 <http://www.biomedcentral.com/1471-2105/12/77/>

If you have any questions regarding the script or this paper, please get in touch.