**PROCESS OVERVEIW**

***Problem Statement:***

The current business state of a Portuguese banking institution, Banco de Portugal, is one of fiscal instability, lost revenue and lowered long-term client engagement (Moro, Cortez, & Rita, 2014). To maintain corporate competitiveness, Banco de Portugal initiated a telemarketing campaign based on the concept that their clientele weren’t depositing as frequently as they had been in the past (Moro, Cortez, & Rita, 2014). The marketing campaign was deployed via direct phone calls to promote the subscription of long-term deposits. With the desire to optimize output and reduce human and financial resources for the campaign, a study was established to determine which targeted clients would most likely respond positively to the telemarketing scheme and subscribe to the long-term deposits.

Distinguishing the success rate of the telemarketing phone calls is of high value as it highlights whether the retail bank in question is able to successfully reach consumers and sell them the campaign. More specifically, which clients are most financially viable to target (in terms of reduced opportunity cost and avoiding financing marketing ploys aimed at the wrong clientele). Marketing campaigns are a typical strategy implemented to enhance business engagement. By using this targeted marketing ploy, the retail banking firm is aiming to meet a desired business state of increased client long-term deposits and consequential financial stability.

This problem has been addressed by past research projects, across various contexts. The eventuated success rate of marketing models are a fundamental component to positively received campaigns. Due to the competitive nature of corporate marketing, data-focused insight and analytical studies are highly prevalent. With specific reference to the data itself, it was retrieved from the original study stemmed from the campaign from 2008-2011 (Moro, Cortez, & Rita, 2014).

Similar studies: \*to be sited and included in referencing, see below\*

<https://ieeexplore.ieee.org/abstract/document/7492828> (Kim, Lee, Jo, & Cho, 2015)

<https://link.springer.com/chapter/10.1007/978-1-349-22313-8_6> (Landreth, 1992)

<https://scholarworks.gsu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1143&context=math_theses> (Tang, 2014)

***Industry/ Domain:***

The banking industry is that in question, with strong focus on marketing operations. With this in mind, the retail banking sector is facing a significant amount of disruption as the digital age is exponentially changes business processes and client behaviour. Specifically the shift towards mobile/ online banking, away from having to go to a physical bank to reap benefits of trade (Lee & Lee, 2020). High competition is prevalent from disruptive start-ups and “neobanks” (Hopkinson & Klarova, 2019) that are changing service expectations - which is only further highlighting that efficiency gains from technology are of elevated value to the consumers.

These factors all motivate the need to adaptively strategize and implement new business processes and market various revenue-stimulating sectors, as is seen in this particular study (Moro, Cortez, & Rita, 2014).

Although the value chain in question centres around fiscal activities (wealth management, financial guidance and asset/liability administration ect.), the study is specifically focused on the success rate of a marketing campaign. Due to the generalizability of such marketing schemes, this project is henceforth highly relevant to multiple industries facing the need to instigate competitive marketing operations.

***Stakeholders:***

The stakeholders in questions are in direct relation to the Banco de Portugal. This covers the official Board of Directors, Board of Auditors, Advisory Board, shareholders and customers of the institution. Each stakeholder is impacted by the success rate of the marketing campaign, promoting long-term deposits. The higher the success rate, the higher the financial investment actions of the banking entity (on product/services with a higher RoR (rate of return). Increasing the banks investment capabilities, increases the services available to clients. For obvious reasons, shareholders will only benefit from increased revenue and stability of long-term deposits, an so forth. From a client perspective, long-term benefits include increased saving capabilities, decreased investment volatility, and increased interest rates.

In line with the project, it can be assumed that clientele wish to gain these competitive rates on their long-term deposits and shareholders will expect to benefit from increased financial flow and business activities to promote the economic success of their institutional foundations.

***Business Question:***

The main business question that requires attention is: *Which existing customers would have a higher probability of responding positively to a long-term deposit marketing campaign?*

In terms of quantifying the value of answering this question, it can be assumed that increased long-term deposits will result in significantly increased financial investment capabilities. The bank can invest the long-term value in various other products that are deemed a higher rate of return.

The required rate of accuracy should be very high as the consequences of incorrect predictability would result in the investment of unsuccessful marketing campaign, increased opportunity cost (missing out on better opportunities/client engagement), reduced revenue and investment capabilities, lowered financial certainty/ increased volatility.

***Data Question:***

From a data science perspective, the question we are attempting to answer is: *Can a Machine Learning model predict which clients are most-likely to successfully respond to telemarketing calls aimed at selling long-term deposits?*

The data required to answer this question involves:

* Survey data of the marketing campaign
* Attributes of the customers on the receiving end of the phone call
* Resulting response/ action of the customer (successful, unsuccessful)

***Data:***

The data was sourced from the UCI Machine Learning Repository (UCI, 2014), based on the aforementioned 2014 study (Moro, Cortez, & Rita, 2014). The data is related with direct marketing campaigns of the Portuguese banking institution, of which were stemmed from direct phone calls. Often, more than one contact to the same client was required, in order to access if the product in question (long-term deposit) would be subscribed ('yes') or not ('no'). There were originally four datasets in question, as follows:

1. bank-additional-full.csv with all examples (41,188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]
2. bank-additional.csv with 10% of the examples (4119), randomly selected from 1), and 20 inputs.
3. bank-full.csv with all examples (45,211) and 17 inputs, ordered by date (older version of this dataset with less inputs).
4. bank.csv with 10% of the examples (4521) and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs).   
   The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g. SVM).

In regards to the reliability of the data, it was established within a highly cited and publicised research paper (Moro, Cortez, & Rita, 2014). The original collection was released publicly for future research and educational purposes, as well as an adaption by the Banco de Portugal that contained revised attributes to accommodate for socio economic factors. Coinciding with the fact that that the collection and study of the data acquired spanned from 2008 to 2014, I can contest that the source is very reliable. With respect to the raw data, the completeness and quality of the figures is varied. There are significant incidents that have unknown variables, as would be expected in the case of telemarketing data. In hindsight, a lot of wrangling will have to be done to get the dataset to a stage that a predictive model can be successfully built upon.

The original data was generated for the founding journal (Moro, Cortez, & Rita, 2014) in the following stages:

1. A Portuguese retail bank was addressed, with data collected from 2008 to 2013 (note the inclusion of the financial crisis).
2. A significant set of 150 features was analysed related with bank client, product and social-economic attributes.
3. A semi-automatic feature selection was initiated in the modelling phase within the study, resulting in a reduced set of 21 attributes.

Although this targeted data was publicised for future studies, the exact figures are not available on an ongoing bases. In the context of the specialized curation of the data with reference to the retail banking case study, this may present as a limitation as future comparisons will not be easily made. However, due to the relatability of the data in correspondence to a multitude of industries, the potential for replication to some degree is implied.

***Attribute Information:***   
*Bank client data:*

1. age (numeric)
2. job : type of job (categorical: 'admin.', 'blue-collar',' entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
3. marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
4. education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
5. default: has credit in default? (categorical: 'no', 'yes', 'unknown')
6. housing: has housing loan? (categorical: 'no', 'yes', 'unknown')
7. loan: has personal loan? (categorical: 'no', 'yes', 'unknown')

*Related with the last contact of the current campaign:*

1. contact: contact communication type (categorical: 'cellular', 'telephone')
2. month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
3. day\_of\_week: last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
4. duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

*Other attributes:*

1. campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
2. pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
3. previous: number of contacts performed before this campaign and for this client (numeric)
4. poutcome: outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')

*Social and economic context attributes:*

1. emp.var.rate: employment variation rate - quarterly indicator (numeric)
2. cons.price.idx: consumer price index - monthly indicator (numeric)
3. cons.conf.idx: consumer confidence index - monthly indicator (numeric)
4. euribor3m: euribor 3 month rate - daily indicator (numeric)
5. nr.employed: number of employees - quarterly indicator (numeric)

*Output variable (desired target):*

1. y - has the client subscribed a term deposit? (binary: 'yes','no')

**DATA SCIENCE PROCESS**

***Data Analysis:***

Abc

***Modelling:***

Abc

***Outcomes:***

Abc

***Implementation:***

Abc

**Results**

***Data Answer:***

Abc

***Business Answer:***

Abc

***Response to Stakeholders:***

Abc

***End-to-End Solution:***

Abc

**References:**

Hopkinson, G., & Klarova, D. (2019, July). How Neobanks’ Business Models Challenge Traditional Banks. *Young Graduate News*.

Kim, K.-H., Lee, C.-S., Jo, S.-M., & Cho, S.-B. (2015, November). Predicting the success of bank telemarketing using deep convolutional neural network. *7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, 314-317.

Landreth, O. (1992). Banco Comercial Português: Putting Marketing at the Forefront. In *European Corporate Strategy* (pp. 84-105). London, United Kingdom: Palgrave.

Lee, S., & Lee, D. (2020). “Untact”: a new customer service strategy in the digital age. *Service Business, 14*, 1-22.

Moro, S., Cortez, P., & Rita, P. (2014, June). A data-driven approach to predict the success of bank telemarketing. *Science Direct, 62*, 22-31.

Tang, H. (2014, December). A Comparison of Two Modeling Techniques in Customer Targeting For Bank Telemarketing. *Scholar Works*.

UCI. (2014, June). *Bank Marketing Data Set.* Retrieved November 2020, from Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing