A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Peter Galvin |
| *Student Number* | SBA24425 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA2 - AI modelling of Bank Deposit marketing |
| *Assessment Due Date* | 2nd January 2025 |
| *Date of Submission* | 2nd January 2025 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Introduction 1](#_Toc186745005)

[Business Strategy 1](#_Toc186745006)

[Objectives and Problem Definition 2](#_Toc186745007)

[Project Scope 2](#_Toc186745008)

[Data Understanding: 2](#_Toc186745009)

[Data Preparation: 2](#_Toc186745010)

[Modeling: 3](#_Toc186745011)

[Evaluation: 3](#_Toc186745012)

[Boundaries: 3](#_Toc186745013)

[Data Sources: 4](#_Toc186745014)

[Timeline: 4](#_Toc186745015)

[Ethical Considerations: 5](#_Toc186745016)

[References: 5](#_Toc186745017)

# Introduction

Conventional banking is facing increasing threats to their business models. Online banks and other Fintechs are attracting customers through personalising products. Digital banks can earn superior profits by avoiding the costs of a physical bank network. (Atiku and Obagbuwa, Feb 2022)

Conventional banks need to amend their marketing and business strategy to face this commercial threat to their profits and customer base.

# Business Strategy

The key strengths of traditional banks are the loyal customer relationships built up over many years. However, if conventional banks do not utilise their customer data to tailor their products then the highly profitable customer segments may be tempted by the digital newcomers.

The global banking sector has witnessed remarkable changes in recent years. The industry is undergoing a digital transformation, changing customer expectations, and disruptive technologies. AI has emerged as a key driver, reshaping the marketing functions. Banks, traditionally seen as conservative institutions, have recognized the urgent need to adapt to the new digital reality.

As customers become increasingly tech-savvy and connected, banks are adopting strategies to remain competitive, relevant, and customer focussed. AI has become essential in reshaping the way banks engage with their customers and create/personalise marketing campaigns. The banking operational model has moved from physical branches to mobile apps, online portals, and digital experiences. Customers now expect seamless, personalized, and on-demand banking services, and banks are embracing AI to meet these expectations.

The aim is to not only attract and retain customers but also to gain insights, optimize resources, reduce operational costs, and foster rapid innovation. (Reddy, 2023)

To improve operational efficiency and enhance profitability traditional banks need to adapt a more customised product marketing strategy. Employing an A.I. model for bank marketing campaigns can help banks reach their commercial and customer goals as follows: (Huang and Rust, 2021)

* An AI model removes the manual process and costs involved in a standard marketing campaign where telephone operators execute blanket sales calls to customers.
* An AI model can easily identify the optimal customers or customer segments leading to tailored product offers which will help increase sales success rates.
* AI models can be used to analyze customer behaviour and test the suitability of products for a particular market segment. (Amado et al., Jan 2018)

The aim of this project is to develop a modern banking AI marketing model that follows an AI driven business strategy. This AI project aims to identify the successful drivers of documented Portuguese term deposit marketing campaign. This data was obtained from a labour intensive blanket marketing campaign by telephone. The AI model aims to identify the customer features in a successful bank term deposit sale. This will avoid the cost of implementing an expensive and labour intensive marketing campaign.

# Objectives and Problem Definition

Develop an AI model to identify the most pertinent customer features that will increase the probability of a successful purchase of a term deposit product.

Developing an AI model will automate many of the conventional marketing processes and subsequently reduce cost. The AI model can also be utilised analyse the optimal sales process. For instance, an intensive and elongated sales process with multiple contacts with the customer may lead to a successful sale but an aggravated customer.

The AI model will be trained using historic Portuguese bank term deposit sale data taken from (May 2008 to November 2010), (“UCI Machine Learning Repository”). The sales data was obtained from direct telephone calls to the bank customers.

# Project Scope

The project will span 2 semesters and will utilise the UCI Portuguese term deposit bank sales data taken from 2008 to 2010. This dataset contains 41k instances of 16 customer and sales process features.

# Data Understanding:

The complete dataset contains 41k rows, however to enable agile model design I have used a random sampled subset of data consisting of 4.5k rows. There are 16 features and one target variable (y for sale, otherwise no).

The dataset contains no missing values and the features contain numeric and categorical data.

Further data subsets may be used in the future to test model accuracy.

# Data Preparation:

For the purposes of implementing machine learning algorithms we need to convert the categorical columns to numerical. I used a process of hot-encoding to break out each categorical data item into its own column feature. This expands the number of features but the process retains all the categorical information.

Plotting the target variable we find almost 500 ‘y’ to sales as opposed to 4000 ‘no’ to sales. The severe data imbalance means the ‘no’ category will overwhelm the modelling process.

I employed the SMOTE process to fabricate synthetic data to balance the target data.

# Modeling:

.The model aims to predict a bank deposit sale (Yes/No) outcome with a dataset containing defined features and target variables. There are various methods available including Regression, Neural network, Knn classification and decision trees. The modelling outcome is to accurately predict a ‘y/n’ answer which is a classification problem. Linear classification or Knn clustering can be used but identifying the underlying drivers is difficult. RandomTree Classification is a method where the underlying driving features can be easily identified.

I have used Keras Random Tree classifier with hyper-parameter training. I have split the dataset into 80%/20% training and testing split. The training data has SMOTE applied to remove the imbalance. GridSearch hyper-parameter training has been used with a selection of nodes, max depth and criterion in order to select the optimal model.

# Evaluation:

Preliminary model results, even after SMOTE are imbalanced. The overwhelming number of ‘no’ target variables seems to have skewed the results. Out of the 98 successful sales in the test data only 18 have been correctly predicted. Further investigation needs to take place into other models to explore the complex underlying relationships. While the overall accuracy score is 90% this includes the ‘no’ predictions which carry little utility. The SMOTE process did not resolve the data imbalance issue which may be due to feature overlap within the data. (Fernandez et al., 2018)

precision recall f1-score support

No 0.91 0.99 0.95 807

Yes 0.60 0.18 0.28 98

accuracy 0.90 905

macro avg 0.75 0.58 0.61 905

weighted avg 0.88 0.90 0.87 905

# Boundaries:

The data is sourced from a Portuguese bank and we must be careful when applying this data to other jurisdictions. The source data is sourced from the 2008 – 2010 timeline and is subject to market dynamics existing for that particular time period. For later timeframes this data may be irrelevant.

The dataset we are using relates to a particular Portuguese bank and term deposit sales. It does not model all banks or all products.

Likewise, the cultural affinity with traditional banks may be relatively strong or weak for Portuguese society compared to Ireland. Additional data in the target country should be obtained before deployment.

# Data Sources:

The UCI Machine Learning repository contains data of direct marketing campaigns for term deposits (phone calls) of a Portuguese banking institution. The data is taken from 2008 to 2010 and includes 16 features and 4.5k instances. (“UCI Machine Learning Repository”)

# Timeline:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Project Timeline (High Level)** | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Data Understanding |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Data Preparation |  |  |  |  |  |  |  |  |  |  |
| - UCI subset I |  |  |  |  |  |  |  |  |  |  |
| - UCI subset II |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Modeling |  |  |  |  |  |  |  |  |  |  |
| - Examine RandomTree |  |  |  |  |  |  |  |  |  |  |
| - Examine SVM |  |  |  |  |  |  |  |  |  |  |
| - Examine Neural Network |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Evaluation |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Document Process |  |  |  |  |  |  |  |  |  |  |
|  | 1 | 2 | 3 | 4 | 5 |  |  | 6 | 7 |  |
|  | **Weeks** | | | | | | | | |  |

The preliminary results indicate a revision of the model data or using a more sophisticated model. Initial indications are the data imbalances are too great and another approach is required. Therefore the new timeline includes further examination of the data sources or new modelling methods.

# Ethical Considerations:

We must be aware that the AI model may yield commercially beneficial but ethically unacceptable outcomes. For example, recommending persistent follow-up to secure a sale may be distressing for the customer and damage the company brand name. (Gonçalves et al., page 3, Feb 2023)

Dealing with financial institution data means we must adhere to GDPR guidelines surrounding customer data security and accuracy. However, data anonymisation in the UCI website reduces this risk. We should also take steps to guard against potential bias in our model such as the unintended exclusion of customer segments from the bank product offers.

# References:

Amado, Alexandra, et al. “Research Trends on Big Data in Marketing: A Text Mining and Topic Modeling Based Literature Analysis.” *European Research on Management and Business Economics*, vol. 24, no. 1, Jan. 2018, pp. 1–7, www.sciencedirect.com/science/article/pii/S2444883417300268, <https://doi.org/10.1016/j.iedeen.2017.06.002>.

Atiku, Sulaiman O., and Ibidun C. Obagbuwa. “Machine Learning Classification Techniques for Detecting the Impact of Human Resources Outcomes on Commercial Banks Performance.” *Applied Computational Intelligence and Soft Computing*, vol. 2021, 21 Sept. 2021, pp. 1–16, <https://doi.org/10.1155/2021/7747907>. Accessed 2 Feb. 2022.

Gonçalves, Ana Rita, et al. “Artificial Intelligence and Its Ethical Implications for Marketing.” *Emerging Science Journal*, vol. 7, no. 2, 14 Feb. 2023, pp. 313–327, www.researchgate.net/publication/368521468\_Artificial\_Intelligence\_and\_Its\_Ethical\_Implications\_for\_Marketing, <https://doi.org/10.28991/esj-2023-07-02-01>.

Huang, Ming-Hui, and Roland T. Rust. “A Strategic Framework for Artificial Intelligence in Marketing.” *Journal of the Academy of Marketing Science*, vol. 49, no. 1, 2021, pp. 30–50. *Springer*, link.springer.com/article/10.1007/s11747-020-00749-9, <https://doi.org/10.1007/s11747-020-00749-9>.

Reddy, S.R.B. (2023). Evaluating The Data Analytics For Finance And Insurance Sectors For Industry 4.0. *Tuijin Jishu/Journal of Propulsion Technology*, [online] 44(4), pp.3871–3877. doi:https://doi.org/10.52783/tjjpt.v44.i4.1559.

Fernandez, Alberto, et al. “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary.” *Journal of Artificial Intelligence Research*, vol. 61, no. 5, 20 Apr. 2018, pp. 863–905, https://doi.org/10.1613/jair.1.11192.

“UCI Machine Learning Repository.” *Archive.ics.uci.edu*, 13 Feb. 2012, archive.ics.uci.edu/dataset/222/bank+marketing.

Github : https://github.com/CCT-Dublin/ca1-capstone-project-proposal-petergalvin-cyber