BANK MARKETING ANALYSIS

Ashel Munemo (am2704)

1. Introduction

This project aims to develop and evaluate several classification models to predict whether a client will subscribe to a term deposit as part of a direct marketing campaign. The dataset used in this analysis is the Bank Marketing dataset, created by created by Paulo Cortez (Univ. Minho) and Sérgio Moro (ISCTE-IUL) in 2012. It contains client demographic information, contact details, socio-economic indicators, and details of previous marketing interactions.

The classification task is binary, with the response variable indicating whether the client subscribed to a term deposit (“yes” or “no”). An initial inspection confirmed that all selected variables were valid, with no missing values, and the dataset contained 45,211 rows and 17 columns.

An extract of the dataset is shown in Table 1.1., which displays the first few rows of the data. The first column represents age, followed by columns describing job type, marital status, education level, and other relevant characteristics.

Table 1.1. Extract of the Data

	age	job	marital	education	default	balance	housing	loan	contact
	<int>	<chr>	<chr>	<chr>	<chr>	<int>	<chr>	<chr>	<chr>
1	58	management	married	tertiary	no	2143	yes	no	unknown
2	44	technician	single	secondary	no	29	yes	no	unknown
3	33	entrepreneur	married	secondary	no	2	yes	yes	unknown
4	47	blue-collar	married	unknown	no	1506	yes	no	unknown
5	33	unknown	single	unknown	no	1	no	no	unknown
6	35	management	married	tertiary	no	231	yes	no	unknown

6 rows | 1-10 of 17 columns

This project implements four different classification approaches to assess how different modelling techniques perform on the same dataset:

1. Logistic Regression
2. Decision Trees
3. Boosting
4. Deep Learning (Neural Network)

The final objective is to determine which model provides the best predictive performance while demonstrating a clear understanding of the modelling techniques, parameter choices, and assumptions involved.

2. Methodology

This study examines the factors influencing whether a client subscribes to a term deposit using the bank marketing dataset. The analysis focuses on understanding how demographic, socio-economic, and previous marketing interaction variables relate to client responses.

To ensure accurate and meaningful insights, data preparation was performed before modelling. The dataset was checked for missing values, and none were found. For models such as boosting and deep learning, the target variable was converted to a numeric binary to ensure compatibility with these algorithms. Additionally, numeric predictors were standardized to ensure faster convergence and stable training of neural networks, while categorical variables were converted to factors for tree- based and GLM models, or one-hot encoded for neural networks.

Exploratory data analysis was used to visualize patterns and relationships between predictors and the target variables. Boxplots, bar charts, and summary tables highlighted distributions and imbalances in the data, helping identify important variables. These visualizations were generated using R's tidyverse package, providing clear and interpretable representations of the data.

This approach ensures that all models are trained on consistent clean data, while exploratory data analysis guides the selection and tuning of models for improved predictive performance.

The following subsections detail the exploratory data analysis and the four classification models used in this study, including the choices made in model fitting and tuning.

2.1. Exploratory data analysis

This section presents the exploratory data analysis, where numerical and categorical predictors are examined in relation to the target variable. The aim is to identify key patterns and preliminary relationships that may influence a client's decision to subscribe to a term deposit.

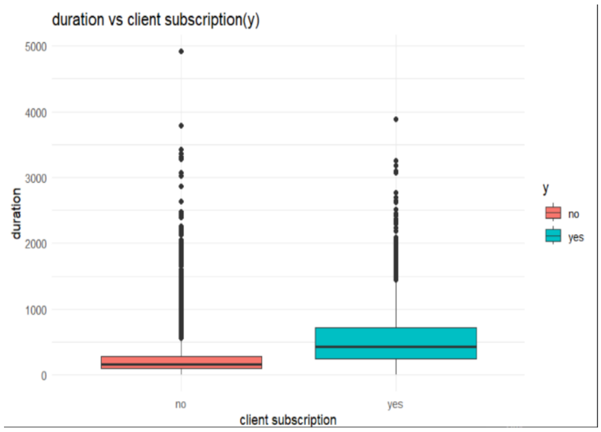
Exploration of categorical variables revealed that the majority of clients did not subscribe across categories such as marital status, job type, and housing and personal loans. Clients with previous successful campaign outcomes were significantly more likely to subscribe, highlighting the importance of campaign history in predicting subscription likelihood.

To support this analysis, a series of figures have been included to illustrate these relationships clearly. Each figure provides insights into how different variables may influence whether a client subscribes to a term deposit or not.

Figure 2.1.0

The figure demonstrates the distribution of clients' balance based on whether they subscribed to a term deposit, along with the results of the corresponding t-test.

```
 T-test for: balance

       Welch Two Sample t-test

data:  balance by y
t = -9.9335, df = 6339.8, p-value < 2.2e-16
alternative hypothesis: true difference in means
between group no and group yes is not equal to 0
95 percent confidence interval:
 -599.3347 -401.7712
sample estimates:
 mean in group no mean in group yes
        1303.715          1804.268
```



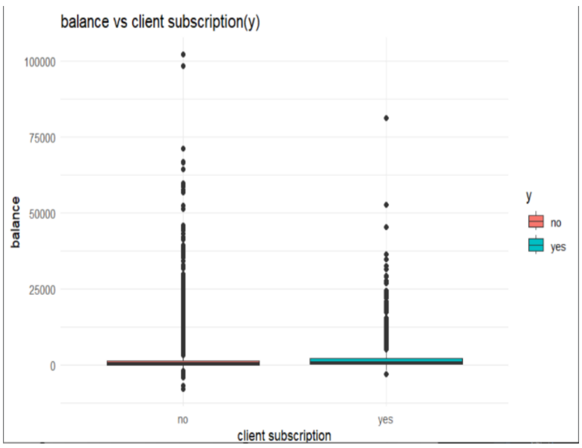
The median balance of clients who did not subscribe to a term deposit is 417 euro, and the median for subscribers is significantly higher, at 733 euros. This suggests that subscribers tend to have a higher average yearly balance than non-subscribers.

Additionally, a t-test was performed to compare the mean balance between clients who subscribed and those who did not. There was strong evidence against the null hypothesis thus we rejected it at 5% significance level ($p < 0.05$), this indicates that a meaningful difference between the two groups of clients exists.

Clients who subscribed had a higher mean balance (1804.27 euros) than those who did not (1303.72 euros). This suggests that account balance is likely important for predicting whether a client will subscribe to a term deposit.

Figure 2.1.1

The plot compares the difference in duration between clients who subscribed to a term deposit and those who did not, along with the results of the corresponding t-test.



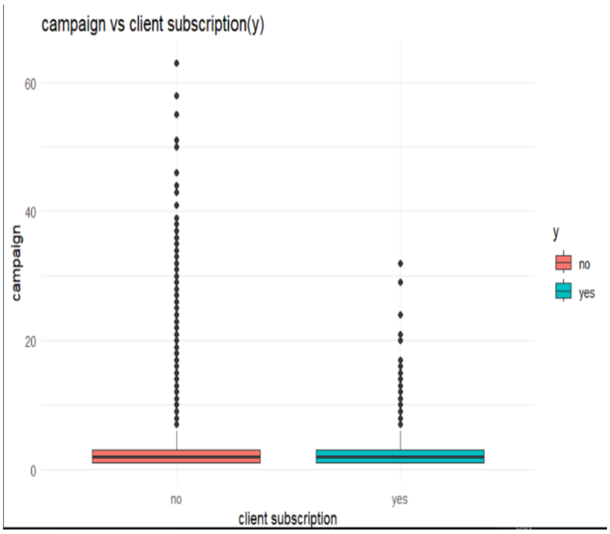

```
 T-test for: duration

       Welch Two Sample t-test

data:  duration by y
t = -57.514, df = 5685.3, p-value < 2.2e-16
alternative hypothesis: true difference in means
between group no and group yes is not equal to 0
95 percent confidence interval:
 -326.8865 -305.3370
sample estimates:
 mean in group no mean in group yes
        221.1828          537.2946
```

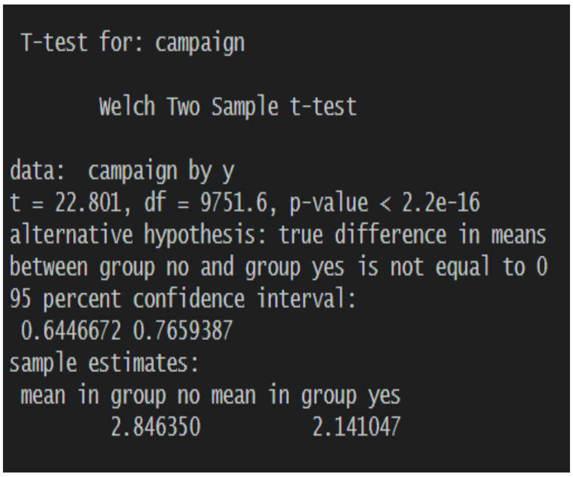
The median of clients who subscribed is noticeably higher, at approximately 426 seconds, while the is median for non-subscribers is approximately 164 seconds. This suggests that clients who subscribed tend to last longer on calls than those who do not subscribe.

A t-test was performed to compare the mean duration between clients who subscribed and those who did not, and strong evidence against the null hypothesis was found thus we rejected the null hypothesis at 5% significance level. Clients who subscribed had a much higher mean duration (537.29) than those who didn't (221.18). This suggests that duration of contact is likely very important for predicting whether a client will subscribe a term deposit.

Figure 2.1.2

The figure below represents the difference in number of contacts performed during a campaign between clients who subscribed to a term deposit and those who did not, along with results of the corresponding t-test.





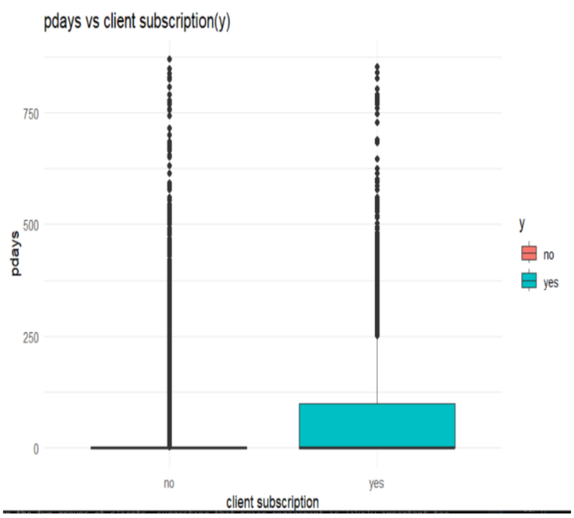
```
T-test for: campaign

       Welch Two Sample t-test

data:  campaign by y
t = 22.801, df = 9751.6, p-value < 2.2e-16
alternative hypothesis: true difference in means
between group no and group yes is not equal to 0
95 percent confidence interval:
 0.6446672 0.7659387
sample estimates:
 mean in group no mean in group yes
         2.846350          2.141047
```

The median was 2 for both groups of clients, those who subscribe to a term deposit and those who did not.

A t-test was also performed to compare the mean number of contacts during a campaign for subscribers and non-subscribers. The results was highly significant ($p < 0.05$), indicating a meaningful difference between the two groups. Clients' who subscribed had a lower mean number of contact during a campaign (2.14) that those who did not (2.85). The suggests that fewer campaign contacts may be associated with higher subscription likelihood.

Figure 2.1.3

The chart compared the number of days that passed by after a client was last contacted from a previous campaign, between clients who subscribed and those who did not, along with the results of the corresponding t-test.




```
T-test for: pdays

       Welch Two Sample t-test

data:  pdays by y
t = -18.943, df = 6251.5, p-value < 2.2e-16
alternative hypothesis: true difference in means
between group no and group yes is not equal to 0
95 percent confidence interval:
 -35.62222 -28.94097
sample estimates:
 mean in group no mean in group yes
         36.42137          68.70297
```

The median of both groups of clients is -1.

A t-test was performed to compare the mean of pdays, and the result was highly significant ($p < 0.05$) indicating that a difference between the two groups exists. Clients who subscribed had a higher mean pdays (68.70) than those who did not (36.42). This suggests that the number of days since last contact is likely very important for predicting whether a client subscribes a term deposit or not.

2.2. Logistic Regression

Logistic regression is a generalised linear model (GLM) used for binary classification. It models the log-odds of subscribing to a term deposit as a linear combination of the predictors. This method allows interpretations through coefficient significance, confidence intervals, and odds ratios, making it suitable for understanding the influence of individual variables.

- **Train/Test split**

The dataset was divided into training and testing subsets to support model building and performance evaluation. Specifically, 70% of the data was allocated to the training set, which is used to develop and fit the predictive model, while the remaining 30% was reserved for the test set, which serves to evaluate how well the model performs to unseen data. This approach helps prevent overfitting by ensuring that the model's accuracy is not solely based on the data it was trained on. By maintaining a fixed random seed, the split remains reproducible, ensuring consistency in results across multiple runs.

- **Model Selection**

A logistic regression model was fitted to predict whether a client subscribes to a term deposit. A stepwise subset selection procedure using Akaike Information Criterion (AIC) was applied to identify the most influential predictors. The initial model included all predictors, and the stepwise process iteratively removed variables that did not improve model performance.

The selection was performed using stepAIC () function, which balances model complexity and goodness of fit.

During model selection, several predictors, were removed as they did not significantly reduce the AIC value. The final model retained a broad set of predictors, these include job, marital status, education, balance, housing, loan, contact, day and month of contact, duration, campaign and outcome of previous marketing campaign.

The final model produced the lowest AIC value, which was 15170.23, indicating a good balance of fit and parsimony.

- Odds ratios , confidence intervals, and parameter significance

The logistic regression results show several factors that significantly influence the likelihood of a client subscribing to a term deposit.

Positive predictors include previous campaign success, which is the strongest driver, with clients who subscribed during past campaigns being dramatically more likely to subscribe again. Longer contact duration and higher average yearly balance also increase subscription odds, though more modestly. Temporals patterns are evident, with subscription peaks in March, June, September, October, and December, while January, May, July, August, and November show lower odds of subscription. Certain client characteristics, such as higher education levels, being retired or being a student, also positively influence whether a client will subscribe to a term deposit.

Negative predictors include existing financial obligations, in essence, clients with housing or personal loans are less likely to subscribe to a bank's term deposit. Job types such as blue-collar, services, housemaid, self-employed, and technician show reduced likelihood of subscription. Married clients are also less likely to subscribe. Contact-related factors have strong effects, for example, unknown and telephone contact both reduce the odds. Finally, campaign factors such as the number of contacts in the current campaign negatively affect client subscription probability, suggesting that repeated contact may lead to client fatigue.

Overall, prior campaign success, education, job type, contact methos, and seasonal timing emerge as the most influential predictors, supported by statistically significant odds ratios and confidence intervals.

- **Threshold tuning**

The fitted model was used to predict probabilities of subscription on the test set. Because the dataset is highly imbalanced, the default classification threshold of 0.5 resulted in poor recall, missing many true subscribers. Various alternative thresholds were evaluated to balance precision and recall, and the following output was obtained:

Cutoff	Accuracy	Precision	Recall	F1- score

0.1	0.815	0.369	0.858	0.517
0.2	0.883	0.496	0.682	0.574
0.3	0.898	0.562	0.552	0.557
0.4	0.904	0.616	0.439	0.512
0.5	0.902	0.638	0.347	0.450

The threshold of 0.2 produced the highest F1-score, indicating the best trade off between precision and recall. This threshold was therefore selected for final classification.

- **Model evaluation**

A confusion matrix was constructed to evaluate prediction performance at the chosen threshold. Metrics such as accuracy, precision, recall, and F1-score were calculated. Although the overall accuracy was high (88.3%), recall was a more relevant metric for capturing actual subscribers due to the imbalanced dataset.

A ROC curve was also generated using predicted probabilities to visually assess model performance. The curve bows strongly toward the top-left corner, indicating good separation between subscribers and non-subscribers. The area under the curve (AUC) was 0.9027, confirming that the model has excellent discriminative ability and correctly identifies subscribers over 90% of the time.

- **Summary of findings**

The logistic regression model successfully identified the key predictors of term deposit subscription. Odds ratios provided interpretable measures of predictor influence, stepwise selection ensured a parsimonious model, and threshold tuning addressed class imbalance. While precision and recall could be further optimized using alternative methods, the final model demonstrates strong predictive performance and provided meaningful insights into factors influencing client subscription.

2.3. Decision Tree

A classification tree model was fitted using all variables to predict whether a client subscribes to a term deposit. During tree construction, only four variables were selected, indicating that they have the strongest discriminatory power. These variables include duration, poutcome, month, and contact.

The resulting tree consists of nine terminal nodes, reflecting a moderately complex model that captures key patterns in the data without being excessively deep. The prominence of duration as the primary split reinforces its importance as a strong predictor, consistent with the findings from the logistic regression model.

Model fit metric show strong performance. The residual mean deviance is 0.4884, suggesting a good overall fit to the data. The misclassification error rate is 10.98%, meaning the tree correctly classifies approximately 89% of all observations. This high accuracy highlights the model's ability to separate subscribing from non-subscribing clients.

The pruning sequence illustrates how model performance changes with complexity. Three candidate subtrees were evaluated, a 10-node tree, a 5-node tree, and a root-only model with one node. The 10-node and 5-node trees both achieve the same deviance value of 3403, whereas collapsing the structure to a single node yields a substantially higher deviance of 3733, indicating a poor fit. Examination of the cost-complexity parameter k further supports this conclusion. The full 10-node tree has k = - infinity, meaning no penalty limits its size, the 5-node tree is selected at k = 0 as the simplest subtree that retains optimal performance, and the 1-node tree emerges only after large penalties are applied. Based on this, the 5-node subtree is the optimal pruning choice under the misclassification criterion.

The final classification tree achieves an accuracy of 88.99%, with a 95% confidence interval of (08845, 0.8951). This accuracy is significantly higher than the no-information rate, demonstrating that the model outperforms a trivial majority-class classifier. The Kappa statistic of 0.437 indicates moderate agreement between predicted and observed outcomes beyond chance.

Class-specific metrics reveal an asymmetric performance pattern. The model is highly effective at identifying clients who do not subscribe, achieving a sensitivity of 94.41% and a positive predictive value of 93.23%. However, classification of the "yes" class is weaker, with specificity of 47.45% and a negative predictive value of 52.55%. The balanced accuracy of 70.93% adjusts for this class imbalance, providing a fairer assessment of performance.

Overall, the decision tree demonstrates strong predictive ability for the majority class ("no") but reduced capability in identifying clients who subscribe to a term deposit. Despite this limitation, the model provides insights into key determinants of client behaviour and offers an interpretable structure that complements more complex models.

2.4. Boosting

The trained dataset was fitted into the boosted model to predict whether a client subscribes to a term deposit. Analysis of variable importance indicates that contact duration is the most influential predictor, followed by previous campaign outcome, suggesting that clients who engaged successfully prior, are likely to subscribe again.

The month of contact also plays an important role, suggesting seasonal or timing effects in client responses. Other variables such as contact, housing loan, age, day of the month, and number of days since last contact contribute modestly to the model. Meanwhile, job type and average yearly balance have very little influence on predictions. Overall, the model relies predominantly on contact duration, previous campaign outcome, and the month of contact to make predictions.

The initial boosted model achieved a sensitivity of 0.325, meaning it correctly identified about one-third of actual subscribers, while maintaining a very high specificity of 0.977. This shows that the model was strong in detecting non-subscribers but missed many true positives. A tuned version with 50 trees, interaction depth 2, and shrinkage 0.1 produced only a slight improvement. Sensitivity increased to 0.337, while specificity remained high at 0.976, and balanced accuracy rose modestly from 0.651 to 0.657. This model was still precise when predicting a subscriber but continued to miss a large proportion of them. To further enhance performance, the number of boosting iterations was increased to 5000 trees using the same tuning parameters. This substantially improved sensitivity to 0.490, balanced accuracy to 0.727, while specificity remained high at 0.962. This model offers the best overall trade-off between identifying subscribers and maintaining predictive accuracy and was therefor selected as the final model.

The ROC curve of the initial model bends strongly toward the top-left corner, indicating good discrimination between subscribers and non-subscribers, supported by an AUC of 0.9043. The tuned model further improved classification performance, confirming better separation between positive and negative classes and aligning with improvements observed in sensitivity and balanced accuracy.

The distribution of predicted probabilities for the initial model ranges from 0.019 to 0.937, with most predictions being low, consistent with the dominance of non-subscribers in the dataset. The tuned model shows a wider range, with extremely low to near-certain probabilities, indicating better separation between likely subscribers and non-subscribers. While the mean predicted probability remains largely unchanged, the maximum predicted probability demonstrates that the tuned model more confidently identifies likely subscribers. Overall, the tuned model provides stronger predictive performance and better distinguishes subscribers from non-subscribers.

2.5. Deep learning

The trained dataset was fitted into a neural network using scaled features, and the model was evaluated using a classification threshold of 0.2 to prioritise detection of subscribers. At this threshold, the model correctly classified 1,215 of 1,566 actual subscribers, achieving a sensitivity of 77.6%, a substantial improvement from the 33.45 at the 0.5 threshold. Specificity remains high at 87.9%, indicating that 10,545 of 11,998 non-subscribers are correctly identified.

Overall, accuracy is 86.7%, which slightly lower than at higher thresholds, reflecting trade-off between correctly identified positives and maintaining majority-class accuracy. The balanced accuracy of 83.7% indicates that the model now performs more equitably across both classes. Threshold tuning clearly demonstrated that the neural network has strong discriminative and that adjusting the decision threshold is an effective method to prioritise detection of the minority positive class without excessively sacrificing accuracy.

The ROC curve for the neural network strongly bows toward the top-left corner, indicating excellent discriminative ability between subscribers and non-subscribers. The AUC of 0.9086 confirms that the model can effectively rank positive cases higher than negative cases, demonstrating that it reliably separates the two classes across all classification thresholds.

3. Results

To evaluate and compare the predictive performance of the models, key metrics including accuracy, balanced accuracy, sensitivity, specificity, and area under the ROC curve were examined in Table 3.1.

Table 3.1.

Model	**AUC**	**Sensitivity (recall)**	**Specificity**	**Precision**	**Accuracy**	**Balanced Accuracy**
Logistic regression	0.9027	0.681	0.910	0.496	0.883	0.795
Decision tree	0.7827	0.944	0.475	0.932	0.890	0.709
Boosting	0.9295	0.490	0.964	0.641	0.909	0.727
Deep learning	0.9086	0.776	0.879	0.455	0.867	0.827

The comparison of the four classification models shows distinct trade-offs between their performance metrics. Boosting achieves the highest AUC, indicating it has the strongest overall ability to discriminate between positive and negative cases. However, it has a relatively low recall, meaning it misses a substantial proportion of actual subscribers, though its high specificity makes it excellent to identify non-subscribers.

Decision tree, on the other hand, has the highest recall and precision, making it effective very effective at capturing subscribers, buts specificity is low, so it produces many false subscribers. Additionally, deep learning shows a strong balance between sensitivity and specificity, resulting in the highest balanced accuracy and solid accuracy, making it the most-well rounded model for both subscriber and non-subscriber detection.

Meanwhile, logistic regression has moderate performance, with recall of 0.681, specificity of 0.910, and balanced accuracy of 0.796, showing it can detect subscribers fairly well while maintaining good accuracy in identifying non-subscribers.

Overall, deep learning offers the best trade-off between detecting subscribers and non-subscribers, boosting is best for non-subscribers, and the decision tree is best at identifying subscribers but at a cost of misclassifying many non-subscribers.

4. Conclusion

To conclude, the analysis reveals several key insights about predicting client subscriptions. Deep learning emerged as the most effective model, providing the best balance between sensitivity and specificity, along with strong overall discrimination and the highest balanced accuracy. This indicates it can reliably identify both potential subscribers and no-subscribers, making it a robust choice for operational deployment.

The EDA revealed important patterns in the data, such as certain client characteristics strongly influencing subscription likelihood, which informed feature selection and model performance. Boosting while achieving the highest AUC tended to miss many actual subscribers, whereas the decision tree captured most subscribers but misclassified many non-subscribers. Logistic regression performed moderately well but was outperformed by deep learning.

Limitations and potential improvements include the relatively imbalanced dataset, which may have affected model sensitivity, and the possibility of improving performance further. Collecting more data on underrepresented clients or exploring new behavioural features could further enhance predictive accuracy.