Classification Analysis of Customer Subscriptions

April Vang[1] apvang@ucdavis.edu
Christopher Li[1] hlcli@ucdavis.edu
Luyang Zhang[1] luyzh@ucdavis.edu
Monica Orme[2] morme@ucdavis.edu

[1]University of California, Davis, Statistics Graduate Studies
[2]University of California, Davis, Biostatistics Graduate Studies

1 Abstract

Machine learning tools have become increasingly attractive as an integral part of business success. We developed a binary classification model that can accurately predict whether a customer will subscribe to a bank's term deposit product after receiving a phone call from a marketer. After reviewing the preliminary results, we focus on the Bayesian regression additive tree (BART) model that was chosen from the four candidate models, including logistic regression, support vector machine, and random forest. We discussed the benefits of BART for this prediction task. We proposed the theory and concepts behind the model and showed the usefulness of BART in revealing important features that marketers can consider for their campaigns. The BART model achieved an area under the curve of 0.802 (95% CI $0.779 - 0.814$) with an accuracy of 90.010%. We showed how social and economic factors impact people's decisions to subscribe to long-term deposits. Lastly, we discussed limitations of the model and future improvements for using machine learning in marketing.

2 Introduction

Machine learning techniques have growing applications in many fields. With the abundance of data and exponentially growing computational power, rich analysis and research have been done for implementing powerful machine learning techniques in industry. Ngai and Wu (2022) reviewed many aspects of machine learning in marketing that are proven to yield many effective marketing decisions. Ma and Sun (2020) proposed that machine learning and big data research in marketing can transform many cases in which traditional quantitative methods were not sufficient. This gives a new perspective of marketing: on top of experience-driven decisions, employers now are more willing to use machine learning to help them unveil breakthroughs in extending marketing techniques, locating potential customers, and increasing sales. The emerging and promising implementation in business motivates us to use some machine learning methods to solve an actual need. Ebrahimi et al.(2020) used a support vector machine to predict consumer purchase behavior. Wu et al.(2021) predicted consumer behavior using random forests in the telecommunications industry.

Thus, we used some real-world data to demonstrate the power of machine learning and how these new techniques can benefit the marketing field. We want to better understand how different factors can influence the success of a bank's telemarketing campaign. We developed some binary classification models to analyze whether a customer is likely to subscribe to a bank's long-term deposit service after receiving a phone call from the telemarketers. We first developed several preliminary models using different machine learning techniques and then proceed to show results focusing on the model that was most preferable. The models predict how likely it is that a potential customer will subscribe to the service given previously collected information.

Therefore, the ultimate goal for developing this model is to motivate bank managers and marketers to be able to identify the customers who are most likely to subscribe to the service. The other goals of this analysis are to understand the characteristics of a customer that have the largest influence on whether they will subscribe to the service and determine how robust the results from the final model are.

3 Dataset and Exploratory Analysis

The data is retrieved from UCI Machine Learning Repository. The source of the data is from a Portuguese banking institution in 2012. It contains 41,188 observations with 19 predictors. The predictors relate to demographics, bank client characteristics, campaign characteristics, and socioeconomic characteristics. For the response, there were more "no" ($n = 36548$ decline to subscribe) than "yes" ($n = 4640$ agree to subscribe). Noting that 10 predictors are categorical with multiple levels, we performed some data transformations so that our data was compatible to be inputted into our various models. We also standardized numerical predictors as needed. The descriptions of each feature can be found in Table 5.

We conducted exploratory data analysis to get to know the relationships between the response variable and different predictors before we fit our models. Tables 3 and 4 give the percentage of yes and no for each level in each of the predictors. They provide us with some preliminary traits of customers who tend to subscribe to the service. Univariate testing for association to subscription status or a difference in means was conducted using chi-square testing or Wilcoxon test.

From Table 3, people who subscribe tend to be older (40.91y vs. 39.91y, p = 0.0161), work in the field of administration and management or are retired, students, or unemployed, and are educated with a university degree or had professional course.

Having a housing loan or personal loan were not found to be significant in its p-values (p = 0.0583 and p = 0.5787 respectively). Marketers generally had more success during the weekdays of Tuesdays, Wednesdays, and Thursdays and more success during the months of March, April, September, October, and December as seen in Table 3 and displayed on bar chart of Figure 1.

Next, let's look at Figure 2 on the social and economic factors stratifed by subscription status. We observe some different distributions for the two subscription statuses. For quarterly average total number of employed citizens, the lower quartile for "no" is almost as high as the mean for "yes", and the mean for employment variation rate for "no" is higher than the upper quartile for "yes". When more people are employed, they are more willing to spend money as they have gotten a stable source of income and less likely to keep their money in deposits. For consumer price index, the lower quartile for "no" is almost as high as the mean for "yes". For consumer confidence index, when people answered "no", the lower quartile is high. This aligned with Bialowolski (2019) study for household consumption–when the consumer confidence index is higher, people are more willing to consume. So when the index is low, it means more people are less willing to spend their money and prefer to deposit it. Therefore, the economy could significantly impact people's decisions to subscribe to long term deposits.

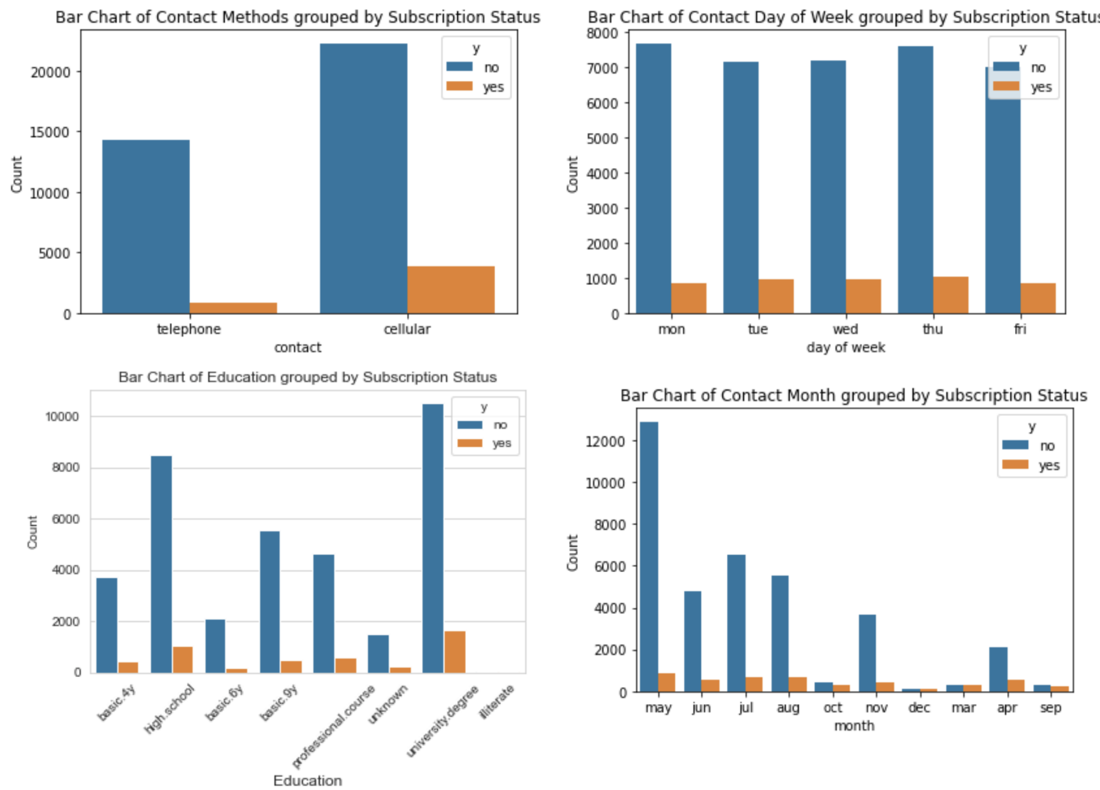

Figure 1: Bar chart for customer traits grouped by subsciption status.

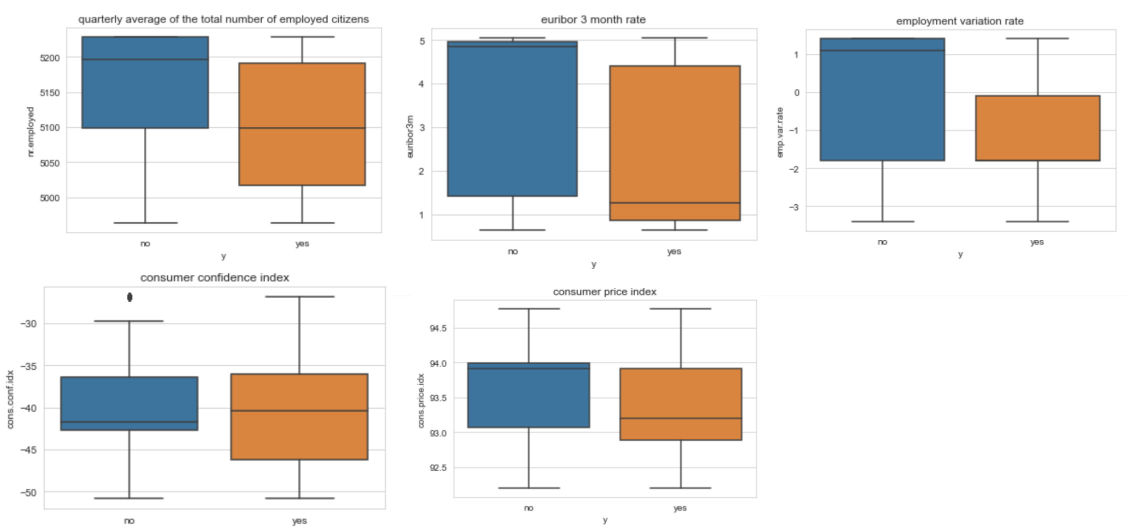

Figure 2: Boxplot for Social-economic indicators grouped by subsciption status.

4 Methods

Four preliminary supervised models were fitted onto the training data, which were Logistic Regression, Support Vector Machine, Random Forest, and Bayesian Additive Regression Tree (BART). Most of these models were fitted onto all predictors, while two Logistic Regression models had been fitted: one using all of features of the dataset and another that used 10 features based on subset selection using the χ^2 statistic for categorical features and the F statistic for quantitative features.

Model Justification. We chose to run these four models based on inspiration from prior marketing literature and for the following reasons. Logistic regression is a solid baseline prediction model to run and is easy interpret, even when we have many categorical variables. Support Vector Machine also made sense to try due to the high-dimensionality of our data. Given the possibility there could be non-linear relationships between our predictors and subscription status, tree-based methods such as random forest and BART emerged as another good option. These tree-based methods were also attractive because of their robustness to noise and outliers. The methodologies are explained in more detail below:

- **Logistic regression** is a parametric model in machine learning that computes the probability for each observation to belong to a class denoted as C (the classes in this analysis are based on subscription status: 0 if a customer does not subscribe and 1 if a customer subscribes) based on a logit function. The probability for an observation to belong to a class is computed as follows:
  P(C = 1 | X = x) = $\frac{e^{\beta^\top x}}{1+e^{\beta^\top x}}$.
  An observation is classified as belonging to class 1 if P(C = 1 | X = x) > θ
  where $\theta \in [0, 1]$, $\vec{\beta} \in \mathbb{R}^{p+1}, x \in \mathbb{R}^{p+1}$.
  In particular, θ is the threshold for deciding how observations are classified, $\vec{\beta}$ represents the vector for the coefficients corresponding to the intercept and each feature, and x represents a single observation from the data matrix X $\in \mathbb{R}^{N\times(p+1)}$
- **Support vector machine (SVM)** is a parametric model that finds an optimal separating hyperplane or a soft margin that gives the most robust prediction for most observations. We tested four kernels for SVM—linear, sigmoid, RBF, and polynomial. If the data is separated by a non-linear boundary, we can use other kernel functions such as polynomial, sigmoid, and RBF to map the data to a higher dimension where the optimal hyperplane or soft margin can be applied. For the final candidate, all features related to jobs were excluded as it gave higher accuracy.
- **Random forest (RF)** is a non-parametric model that combines the results from multiple decision trees, and is an extension of bagging. In bagging, multiple trees are fit on bootstrapped data and predicted probabilities of each class are averaged across all trees. Random forest differs from bagging in that each individual tree is fit by only considering a random subset of features when looking for the best split at each node. In this way, random forest helps to decorrelates the trees in the forest. Furthermore, the maximum depth of the random forest is tuned using cross validation in order to prevent overfitting.
- **Bayesian additive regression trees (BART)** is a non-parametric method that fuses the Bayesian approach with sum-of-trees to compute posterior probabilities and make inferences. Each tree explains a portion of the data instead of the entire data with one tree. A Bayesian backfitting MCMC iteratively constructs and fits successive residuals that combines methods from both Gibbs sampling and Metropolis Hastings. Prior parameters the use of tree perturbation helps to prevent BART from overfitting. BART determines nonlinear effects automatically without much specifications from users.

5 Model Evaluation

5.1 Model Evaluation Methods

The original dataset (N = 41,188) was split into two non-overlapping datasets:

- 80% of the customers in the original dataset were randomly selected to be a part of the training dataset (n = 32,950). This training dataset was used to build the classification models that were discussed in the methods section.

- 20% of the remaining customers in the original dataset were a part of the testing dataset (n = 8,238). This dataset was reserved for predicting the customers' subscription status to a bank's term deposit using the constructed models that were built using the training dataset.

Classification metrics (classification rate, AUC, sensitivity, specificity, and the F1 score) were obtained for each model. The classification rates and the AUC values from each model were used to compare and evaluate the models in order to determine which model provides the optimal predictions in determining a customer's subscription status. The remaining classification metrics were used to compare how well the models detect customers who actually subscribe to a bank's term deposit and those who do not. The interpretations and the usefulness of these classification metrics will be discussed in the following section.

5.2 Model Evaluation Metrics

A few classification metrics have been used to compare the performance of the previously introduced models on the testing data. In particular, the classification rate and the AUC were used to determine the optimal model. The specificity, sensitivity, and F1 score were used to assist with understanding how well the model predicts a particular response. As model comparison metrics are important to understand, we provide their definition and explanation notations with this example. We will use a confusion matrix (shown below) to explain these:

		True Label	
		Positive (+)	Negative (-)
Predicted Label	Positive (+)	TP	FP
	Negative (-)	FN	TN

Here, a positive result (+) refers to customers who subscribe to a bank's term deposit, whereas a negative result (-) refers to customers who do not subscribe to that.

Each of the cells in the confusion matrix can be interpreted as follows (Hastie, et al., 2017):

- TP: The number of customers that are correctly predicted as subscribing to a bank's term deposit.
- TN: The number of customers that are correctly predicted as not subscribing to a bank's term deposit.
- FN: The number of customers that are misclassified as not subscribing to a bank's term deposit when they actually subscribed.
- FP: The number of customers that are misclassified as subscribing to a bank's term deposit when they actually did not subscribe.

Using the notations from the confusion matrix, the notations and interpretations of the classification metrics can be understood in the following way:

- Classification Rate (Accuracy): The percentage of the time that the true subscription status of a customer is predicted correctly.

$$\text{Accuracy} = \frac{(TP+TN)}{N} \text{ where N = TN+TP+FN+TN}$$

- AUC: Metric of how well the classifier can distinguish customers who do and do not subscribe to a bank's term deposit across various thresholds, which is measured by computing the area under the ROC curve. This also measures the classification performance and still accounts for both the true positive rates (TPR) and false positive rates (FPR).
- Specificity (True negative rate): The probability of a negative test result, conditioned on the individual truly being negative. In context, the specificity refers to the estimated probability that a customer is predicted to not subscribe to a bank's term deposit given that they actually did not subscribe to a bank's term deposit.

$$\text{Specificity} = \frac{TN}{(TN+FP)}$$

- Sensitivity (True positive rate or the recall rate): The probability of a positive test result, conditioned on the individual truly being positive. In context, the sensitivity refers to the estimated probability that a customer is predicted to subscribe to a bank's term deposit given that they actually subscribed to a bank's term deposit.

$$\text{Sensitivity} = \frac{TP}{(TP+FN)}$$

- Precision: Percentage of customers that actually subscribe to a bank's term deposit given that they are predicted to subscribe (Precision-Recall, 2023).

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

- F1 Score: A weighted average of the precision and recall. A high F1 score is desirable.

$$\text{F1} = \frac{2(Precision*Sensitivity)}{(Precision+Sensitivity)}$$

6 Preliminary Results

In order to determine the model that provides the most accurate predictions, the AUC values and classification rates for each model were compared. The model with the highest AUC and classification rates best satisfy these objectives. We are interested in these particular metrics because both of these measure prediction accuracy and the AUC accounts for the true and false positive rates. The sensitivity, specificity, and F1 scores were also obtained to provide information about how well the models detect customers who actually did and did not subscribe to a bank's term deposit, but were not the primary factors in deciding the final model.

Model (n features)	Accuracy (%)	AUC	Sensitivity	Specificity	Precision	F1-Score
Log. Reg. (10)	89.718%	0.786	0.228	0.985	0.66	0.339
Log. Reg. (all)	89.7%	0.794	0.226	0.984	0.654	0.336
SVM (featres for jobs excluded)	88.6%	0.76	0.032	0.888	0.714	0.06
Random Forest (all)	89.7%	0.804	0.215	0.986	0.668	0.326
BART (all features)	90.010%	0.802	0.315	0.977	0.637	0.422

Table 1: Classification Metrics from the Preliminary Models on the Test Data

Each of the preliminary models yielded high classification rates (at least 88%), so most of these models often predicted the customer's subscription status correctly. The AUC values for these models ranged from 0.76 to 0.804, which indicates that the classifiers of these models can distinguish customers who do and do not subscribe to a bank's term deposit across various thresholds perform moderately well overall. In particular, the BART model yielded the highest classification rate (90.01%) and an AUC of 0.802. While its AUC is the second highest among the preliminary models, this model has been deemed the final model, the one that provides the most accurate predictions since its classification rate was the highest and its AUC was 0.002 points away from the highest AUC value from the Random Forest model. The BART model performs the best at detecting customers that do subscribe to a bank's term deposit (Sensitivity = 31.5%) compared to the other attempted models. Hence, the BART model has been considered the final model of interest. It is worth noting that additional analysis under the random forest model and logistic regression (not shown) reveal similar results, but we focus on the BART model.

Support vector machine underperformed for this dataset, and this could be because the imbalanced number of "yes" and "no" in the response variable. Akbani et al.(2004) addressed the problem that SVM does not perform as well if one level in the response outnumbers the other one. In our data, 36,548 decline the service whereas only 4,640 customers subscribed. This imbalanced number could deviate the SVM classifer and return a lower accuracy in validation. This resulted in high false positive and low sensitivity and F-1 score in Table 1. SVM tends to misclassify many observations into positive even though the accuracy is not significantly low. Thus, SVM is not an optimal classifier both conceptually and empirically.

Since BART has been deemed as the final model of interest, further analysis has been conducted using this model in order to address the remaining project objectives. In particular, the most important features from the model and a sensitivity analysis on this model will be conducted in order to learn more about how different factors can influence the success of a bank's telemarketing campaign and determine if the results from this model are robust.

7 Proposed Method: BART Algorithm and Robustness

Assumptions. BART is a flexible model that can handle complex confounding and main effects relationship with the outcome. We assume $\epsilon \sim N(0, \sigma^2)$ and a non-parametric approach.

Model and Method. Let the output Y be $Y = f(x) + \epsilon$ and we use p dimensional predictor vectors $x = (x_1, ..., x_p)$. We approximate $f(x) = E(Y|x)$ using a sum of m regression trees. The BART model consists of two components (1) a sum-of-trees model and (2) a regularization prior on the parameters of that model.

For the sum-of-trees model. Let $j = 1, ..., m$ for m total trees and T_j denote a binary tree consisting of interior node decision rules and a set of terminal nodes, and let $M = \{\mu_1, \mu_2, ...\mu_b\}$ denote a set of parameter values associated with each of the b number of terminal nodes in T_j.

Let the classification probability $p(x) = P[Y = 1|x] = \Phi[G(x)]$ using the standard normal cumulative distribution function and $G(x) = \sum_{j=1}^{m} g(x; T_j, M_j)$. We have that $g(x; T_j, M_j)$ is the function that assigns $\mu_{ij} \in M_j$ to x as it searches through the sequence of decision rules from top to bottom and x is assigned a single μ_{ij} value. Next, the sum-of-trees would be the addition of the resulting μ_{ij} assigned to each x. In practice, $g(x; T_j, M_j)$ is unknown, and we therefore need prior distributions. BART defines two priors under classification which is (1) a prior for the a tree T_j which is the probability for the node to split and its likeliness to allow for convergence and (2) a prior for the parameter μ_{ij} chosen in a specific way to yield a 95% prior probability that $E(Y|x)$ is in the interval (y_{min}, y_{max}). BART assumes $\sigma = 1$ under classification problems.

Algorithm. Next, the main algorithm BART uses is the Bayesian backfitting Monte Carlo Markov Chain (MCMC) that iteratively constructs and fits successive residuals. Methods from both Gibbs sampling and Metropolis Hastings are used and the complexity in this MCMC algorithm are due to the unique structure of the joint posterior distribution that needs to be estimated.

Within each iteration t of the MCMC algorithm, draws for each tree's parameters are made tree-by-tree, conditioned on all other trees' parameters which results in Gibbs sampling. Let us denote the structure and terminal node parameters of all trees except for the jth tree as $T_{(j)}$ and $M_{(j)}$ respectively. In each iteration, there will be m number of draws (cycling through all m trees), where each draw will be from the conditional distribution involving $T_{(j)}$ and $M_{(j)}$.

The parameters of a given tree T_j consist of the splitting rules at each node, which includes which predictors are used for splitting and what the splitting threshold value is. The parameters of T_j are collectively drawn from a proposal density, since Metropolis Hastings is being used. The new draw of T_j^t at iteration (t) comes from a proposal density containing the specific set of possible trees that can be derived from applying a variety of changes to the tree T_j^{t-1}. These changes include growing the tree (randomly picking a terminal node and splitting it based on splitting rule set in the prior), pruning a tree (collapsing two terminal nodes into its parent node), altering the splitting rule of an internal node, or swapping the splitting rules between a parent node and a child node.

After running this MCMC algorithm on a training dataset, one can use this model for prediction on a test dataset. The first step is to identify how many burn-in draws to use, so that all the after burn-in draws can be used as samples from the posterior distribution. Then for each post burn-in draw, one can use the parameters from this draw (which define the structure of all m trees of that draw) to predict the response variable ($\hat{y}^{test}$) for a given set of predictor variable values (x^{test}). Lastly, averaging the predicted response variable values across all post burn-in draws will provide a final prediction, which will be an estimate for the posterior mean $E(\hat{y}^{test}|x^{test}, y^{train}, x^{train})$.

7.1 BART Analysis and Robustness

We used 200 trees for the BART model. Figure 3 details the convergence diagnostics and features for the MCMC iterations. Figure 3A shows the proportion of Metropolis-Hastings steps that is accepted for each Gibbs sample. We observe a range between 0.2 and 0.5 acceptance probability. Figure 3B and 3C displays the average number of nodes or average tree depth respectively for each tree according to the Gibbs sample for the post burn-in. We observe that the average number of nodes is roughly 4 and average number of tree depth is roughly 1.5 as indicated by the blue line.

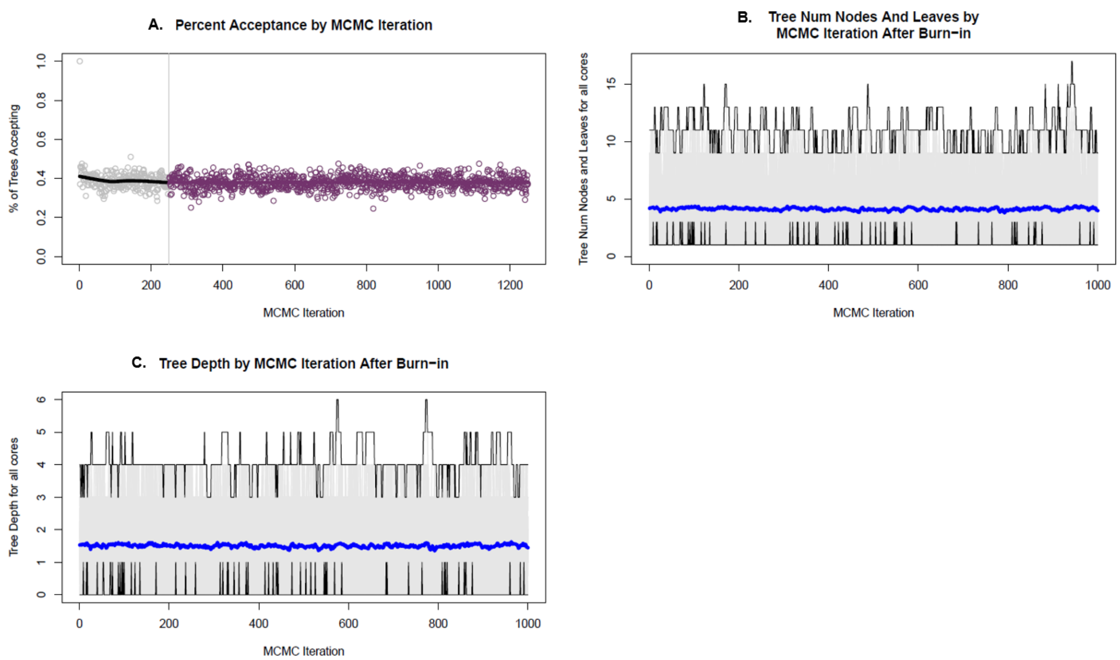

Figure 3: MCMC Iteration Characteristics

Next, to ensure a robust model, we consider testing BART under slight parameter changes by reviewing results under different number of trees and at different seeds. We observe the following results in Table 2.

Seed	Time (min)	Trees	Accuracy	AUC	95% AUC CI	Sens.	Spec.	Prec.	F1
1	1.773599	15	89.378%	0.7977	0.7811 – 0.8142	0.266	0.976	0.590	0.366
208	1.528857	15	89.791%	0.7917	0.7745 – 0.809	0.301	0.976	0.620	0.406
1	17.69184	50	89.342%	0.7996	0.783 – 0.8162	0.262	0.976	0.587	0.362
208	16.85052	50	89.694%	0.7944	0.7773 – 0.8114	0.300	0.975	0.610	0.402
1	61.32122	200	**90.010%**	**0.802**	**0.7791 – 0.8142**	**0.315**	**0.977**	**0.637**	**0.422**
208	58.25681	200	89.755%	0.7909	0.7718 – 0.8100	0.263	0.981	0.638	0.372

Table 2: BART Performance

The amount of time taken to run BART for all conditions in Table 2 is between 2 to 61 minutes per model. We observe that this model has similar performance when the number of trees or seed changes since the classification metrics do not vary much with these changes. This demonstrates a robust model that is not sensitive to small parameter changes which makes it a desirable in a model.

BART vs. Logistic Regression and Random Forest. We also highlight that BART has similar performance to Logistic Regression and Random Forest. To provide intuition on why this may be, we can think of a regression tree that goes from the top down checking conditions and view it as an analysis of variance (ANOVA) model containing indicator functions with main effects and multi-way interaction effects. Each

MCMC iteration can produce a different stepwise ANOVA model that we sum over, and summing results in a stepwise approach as seen below.

$$y = \mu_1 \mathbb{1}_{\text{tree-condition1}} + \mu_2 \mathbb{1}_{\text{tree-condition2}} + \mu_3 \mathbb{1}_{\text{tree-condition3}} + \dots + \mu_b \mathbb{1}_{\text{tree-condition-b}}$$

In a sense, it combines both the tree-based aspect, and a regression aspect as seen below with indicator functions that satisfies the set of interior node decision rules. We can explicitly see the main effect and multiple-way interactions and this makes BART a flexible model.

8 Key Findings

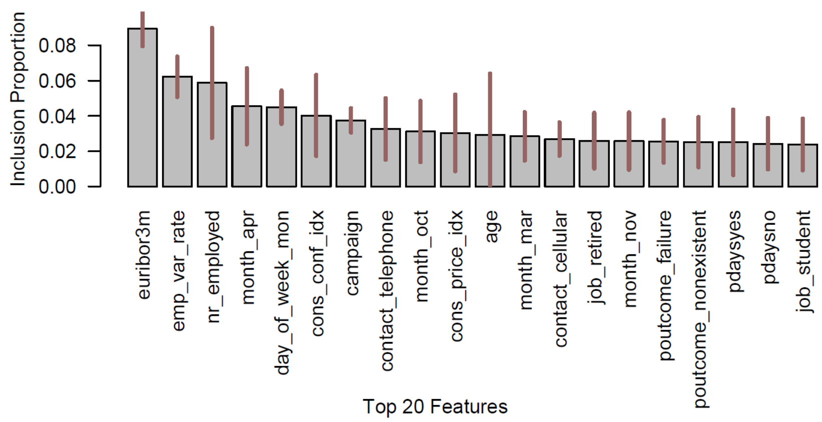

Figure 4: Top 20 Most Important Features

Based on Figure 4, the feature importance chart from the BART model, one can observe the features that are most important in helping to predict whether or not a customer will subscribe to a term deposit product. Feature importance is measured by the inclusion proportion, which represents the "the proportion of times a variable was selected over all posterior MCMC samples in the sum of trees model" (Hernandez et al. 2018). Three economic indicators: the daily three month Euribor rate, the quarterly employment variation rate, and the quarterly average of the number of employed citizens in Portugal have the highest inclusion proportion. The daily three month Euribor rate is a commonly used index for interest rates across Europe, and many European banks set their deposit interest rates to follow the three month Euribor rate (Moro et al., 2014).

It is common economic knowledge that central banks will lower interest rates to stimulate economic growth and raise interest rates to slow down economic growth. Based on Table 4 and the strong importance of the Euribor rate demonstrated in Figure 4, it is evident that on average, customers subscribed to term deposits more often when interest rates were lower ($p < 0.0001$). At first, this result may seem counter intuititive, because with a lower interest rate, customers would be making less money off their term deposits. However, it is important to notice that this data was collected in the midst of the 2008 financial crisis. Therefore, low interest rates may be correlated with the time during the financial crisis when economic conditions were already worsening and the European Central Bank felt the need to try and stimulate the economy more by lowering interest rates. At this time, citizens may have felt the need to save more money by putting their money into term deposits, due to the tough economy (Moro et al., 2014). Therefore, the importance of the Euribor rate in predicting customer subscription to term deposits could be a consequence of the Euribor rate being a clear representation of the state of the economy.

In our earlier exploratory analysis, arguments are provided for why employment information may be significant in predicting subscription to term deposits. It is plausible to think that the importance of the employment variation rate and the number of employed citizens in Portugal in our model could also be rationalized using logic similar to that used for the Euribor rate above. Table 4 demonstrates that on average, customers subscribed to term deposits more often when the number of employed citizens was lower and decreasing. The employment rate was lower during the depths of the financial crisis when the economy was struggling, and it would make sense that customers were more inclined to saving money during this time. This would explain why the number of employed citizens and the employment variation rate would be important predictors in our analysis. We can also see that some variables indicating the month the customer was contacted were also important. This makes sense given that over the time period of the data, the economy was worse in some months and better in other months.

Over the 10-month period, most calls 33.43% were made in May but relative to other months, fewer people subscribed (19.09% vs. 35.25%). We observed more successes during the months of March (accounting for 5.95% of all subscriptions vs. 0.74% of all non-subscriptions), April (11.62% vs. 5.73%), September (5.52% vs. 0.86%), and October (6.79% vs. 1.10%) with more emphasis during March, April, and October as they were among the top twelve most important features revealed by BART. The month of November reached 15th most important and inspection of Table 4 reveals that telemarketing campaigns observed less successes during this time in November (8.97% vs. 10.08%).

Calls throughout the weekdays are roughly equally dispersed from Monday to Friday (20% per day), however for future telemarketing campaigns, we suggest to instead call more during the weekdays of Tuesday, Wednesday, and Thursday and less on Monday and Friday. Calling on Monday was ranked 5th most important and pairing this information with Table 4 further supports our suggestion to call less on Monday since it accounted for a higher percentage of unsuccessful campaigns compared to successful campaigns (18.25% vs. 20.98%).

Another notable finding is that the BART model feature importance analysis finds that prior customer contacts before the current campaign are useful in predicting customer subscriptions. The 'pdays' variable, which we dichotomized, indicates whether or not a client was contacted in the prior campaign. The 'poutcome' variable represents whether a prior campaign was successful (whether the client subscribed to the product). Table 4 demonstrates that a higher percentage of customers were contacted in a prior campaign for the the customers who subscribed to the term deposit product, compared to those who didn't subscribe. Similarly, a higher percentage of customers were successful conversions (subscribed) in a prior campaign for the the customers who subscribed to the term deposit product, compared to those who didn't subscribe. This illustrates that prior campaign activity has a positive influence on current campaign success.

Education did not reach the top 20 most important, but we noticed a trend that individuals with a high school degree or lower are less likely to subscribe compared to individuals with a professional course or university degree who will more likely subscribe ($p < 0.0001$). If telemarketing campaigns have the resources, we suggest they promote the importance of subscribing and its benefits to less educated folks by making them more aware to help target this group to subscribe more.

9 Conclusion and Discussion

The BART model was selected as the model of interest since it most accurately predicted a customer's subscription status to a term deposit product compared to the other tested models and had a higher sensitivity rate. After assessing the model's most important features, it has been shown that the employment variation rate, the quarterly average of the total number of employed citizens, and the daily three month Euribor rate had the largest influence on whether or not a customer will subscribe. Furthermore, this model is robust to changes in the hyperparameters and different seed settings since the classification metrics do not change drastically when these settings are adjusted (Table 2). It is important to note that while this model has desirable characteristics and addresses our project objectives, its prediction performance was only marginally better than most of the other tested models from the preliminary results.

Based on the analysis of feature importance from the BART model, it is clear that the state of the economy is an important predictor for if a customer will subscribe to bank deposits. Additionally, factors such as whether a customer had been contacted in prior marketing campaigns and the success of those prior contacts was also important. Therefore, there are two clear recommendations for bank marketing campaigns in the future. One is that it would be advantageous to scale up marketing campaigns when the economy is struggling, since customers may be more likely to subscribe during these periods. This makes sense since term deposits can be seen as a saving mechanism, and saving becomes more important during economic downturns. Another key recommendation is to target customers who have been contacted in prior campaigns, especially those who subscribed to products in prior campaigns. Our analysis indicates that these repeat customers are more likely to subscribe to the term deposit product. This makes intuitive sense given the idea that past customers are more likely to subscribe to more products in the future.

Although BART has been shown to be a robust and highly accurate model, there are still some caveats to this analysis. The overall accuracy was high for this model (90.01%), but by using the default probability threshold, the model could only detect customers who subscribed to a bank's term deposit at a rate of 31.5% . However, this low sensitivity could have been improved by changing the threshold of the model to more accurately predict customers that subscribe. Ultimately, one could tune the threshold of this model to achieve the specific goals desired. Banks who have more abundant financial resources to carry out marketing campaigns could identify a wider net of potential customers by leaving the probability threshold lower. On the other hand, a bank that wants to be very selective in its marketing campaign could raise the threshold to predict fewer candidate customers that would subscribe, which would help decrease the number of false positives. Furthermore, the model selection process could have been more informative by comparing how well the each of the models' assumptions were satisfied instead of primarily considering how accurate the predictions were across the preliminary models.

10 References

Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. (1998). "Bayesian CART model search (with discussion and a rejoinder by the authors)," *Journal of the American Statistical Association*, 93, 935–960.

Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. (2010). "BART: Bayesian additive regression trees," *Annals of Applied Statistics*, 4(1), 266–298.

Hastie, T., Friedman, J., and Tisbshirani, R. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. - Chapter 9. Additive Models, Trees, and Related Methods* Springer.

Hernandez, Belinda, Adrian E. Raftery, Stephen R. Pennington, and Andrew C. Parnell. (2018). "Bayesian Additive Regression Trees using Bayesian Model Averaging," *Statistics and Computing*, 28(4), 869–890.

Moro, Sergio, Paulo Cortez, Paulo Rita. (2014). "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems*, 62, 22–31.

Ma, Liye., Baohong Sun.(2020). "Machine learning and AI in marketing – Connecting computing power to human insights," *International Journal of Research in Marketing*, 37(3), 481-504.

Pejman, Ebrahimi., Aidin Salamzadeh, Maryam Soleimani, Seyed Mohammad Khansari, Hadi Zarea, and Maria Fekete-Farkas. (2022). "Startups and Consumer Purchase Behavior: Application of Support Vector Machine Algorithm" *Big Data and Cognitive Computing*, 6(2), `https://doi.org/10.3390/bdcc6020034`.

Precision-recall. scikit learn. (2023). `https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html#precision-recall`

S. Wu, W. -C. Yau, T. -S. Ong and S. -C. Chong. (2021). "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," *IEEE Access*, 9, 62118-62136, doi: 10.1109/ACCESS.2021.3073776.

Akbani, R., Kwek, S., Japkowicz, N. (2004). "Applying Support Vector Machines to Imbalanced Datasets," *Machine Learning: ECML 2004–Lecture Notes in Computer Science*, 3201, `https://doi.org/10.1007/978-3-540-30115-8_7`.

Bialowolski, Piotr.(2019). "Economic sentiment as a driver for household financial behavior," *Journal of Behavioral and Experimental Economics*, 80, 59-66, `https://doi.org/10.1016/j.socec.2019.03.006`.

11 Appendix

11.1 Code

The majority of the code has been completed in Python using .ipynb files. The code for building and running the BART model was conducted in R since there is currently no accessible module for using BART for classification in Python.

The code for this project can be found in the following Github repository: Github Repository

11.2 Supplementary Tables

Table 3: Demographics by Subscription

Variable		Overall (n=41188)	No (n=36548)	Yes (n=4640)	P-value
Demographics					
Age		40.02 (10.42)	39.91 (9.9)	40.91 (13.84)	0.0161
Job					<0.0001
	administration	10422 (25.3%)	9070 (24.82%)	1352 (29.14%)	
	blue-collar	9254 (22.47%)	8616 (23.57%)	638 (13.75%)	
	entrepreneur	1456 (3.54%)	1332 (3.64%)	124 (2.67%)	
	housemaid	1060 (2.57%)	954 (2.61%)	106 (2.28%)	
	management	2924 (7.1%)	2596 (7.1%)	328 (7.07%)	
	retired	1720 (4.18%)	1286 (3.52%)	434 (9.35%)	
	self-employed	1421 (3.45%)	1272 (3.48%)	149 (3.21%)	
	services	3969 (9.64%)	3646 (9.98%)	323 (6.96%)	
	student	875 (2.12%)	600 (1.64%)	275 (5.93%)	
	technician	6743 (16.37%)	6013 (16.45%)	730 (15.73%)	
	unemployed	1014 (2.46%)	870 (2.38%)	144 (3.1%)	
	unknown	330 (0.8%)	293 (0.8%)	37 (0.8%)	
Marital Status					<0.0001
	single	11568 (28.09%)	9948 (27.22%)	1620 (34.91%)	
	married	24928 (60.52%)	22396 (61.28%)	2532 (54.57%)	
	divorced	4612 (11.2%)	4136 (11.32%)	476 (10.26%)	
	unknown	80 (0.19%)	68 (0.19%)	12 (0.26%)	
Education					<0.0001
	basic 4 years	4176 (10.14%)	3748 (10.26%)	428 (9.22%)	
	basic 6 years	2292 (5.56%)	2104 (5.76%)	188 (4.05%)	
	basic 9 years	6045 (14.68%)	5572 (15.25%)	473 (10.19%)	
	high school	9515 (23.1%)	8484 (23.21%)	1031 (22.22%)	
	illiterate	18 (0.04%)	14 (0.04%)	4 (0.09%)	
	professional course	5243 (12.73%)	4648 (12.72%)	595 (12.82%)	
	university degree	12168 (29.54%)	10498 (28.72%)	1670 (35.99%)	
	unknown	1731 (4.2%)	1480 (4.05%)	251 (5.41%)	

Table 4: Bank, Campaign, and Socioeconomic Characteristics by Subscription

Variable		Overall (n=41188)	No (n=36548)	Yes (n=4640)	P-value
Bank Client Characteristics					
Has credit in default					<0.0001
	no	32588 (79.12%)	28391 (77.68%)	4197 (90.45%)	
	yes	3 (0.01%)	3 (0.01%)	0 (0%)	
	unknown	8597 (20.87%)	8154 (22.31%)	443 (9.55%)	
Has housing loan					0.0583
	no	18622 (45.21%)	16596 (45.41%)	2026 (43.66%)	
	yes	21576 (52.38%)	19069 (52.18%)	2507 (54.03%)	
	unknown	990 (2.4%)	883 (2.42%)	107 (2.31%)	
Has personal loan					0.5787
	no	33950 (82.43%)	30100 (82.36%)	3850 (82.97%)	
	yes	6248 (15.17%)	5565 (15.23%)	683 (14.72%)	
	unknown	990 (2.4%)	883 (2.42%)	107 (2.31%)	
Contact method					<0.0001
	cellular	26144 (63.47%)	22291 (60.99%)	3853 (83.04%)	
	telephone	15044 (36.53%)	14257 (39.01%)	787 (16.96%)	
Last contact month					<0.0001
	mar	546 (1.33%)	270 (0.74%)	276 (5.95%)	
	apr	2632 (6.39%)	2093 (5.73%)	539 (11.62%)	
	may	13769 (33.43%)	12883 (35.25%)	886 (19.09%)	
	jun	5318 (12.91%)	4759 (13.02%)	559 (12.05%)	
	jul	7174 (17.42%)	6525 (17.85%)	649 (13.99%)	
	aug	6178 (15%)	5523 (15.11%)	655 (14.12%)	
	sep	570 (1.38%)	314 (0.86%)	256 (5.52%)	
	oct	718 (1.74%)	403 (1.1%)	315 (6.79%)	
	nov	4101 (9.96%)	3685 (10.08%)	416 (8.97%)	
	dec	182 (0.44%)	93 (0.25%)	89 (1.92%)	
Last contact weekday					<0.0001
	mon	8514 (20.67%)	7667 (20.98%)	847 (18.25%)	
	tue	8090 (19.64%)	7137 (19.53%)	953 (20.54%)	
	wed	8134 (19.75%)	7185 (19.66%)	949 (20.45%)	
	thu	8623 (20.94%)	7578 (20.73%)	1045 (22.52%)	
	fri	7827 (19%)	6981 (19.1%)	846 (18.23%)	
Last contact duration (sec)		258.29 (259.28)	220.84 (207.1)	553.19 (401.17)	<0.0001
Campaign Characteristics					
Previously contacted by another campaign					<0.0001
	no	39673 (96.32%)	36000 (98.5%)	3673 (79.16%)	
	yes	1515 (3.68%)	548 (1.5%)	967 (20.84%)	
No. of contacts for this campaign		2.57 (2.77)	2.63 (2.87)	2.05 (1.67)	<0.0001
No. of contacts before this campaign		0.17 (0.49)	0.13 (0.41)	0.49 (0.86)	<0.0001
Outcome of previous marketing campaign					<0.0001
	failure	4252 (10.32%)	3647 (9.98%)	605 (13.04%)	
	success	1373 (3.33%)	479 (1.31%)	894 (19.27%)	
	nonexistent	35563 (86.34%)	32422 (88.71%)	3141 (67.69%)	
Social and Economic Characteristics					
Emplyment variation rate (quarterly)		0.08 (1.57)	0.25 (1.48)	-1.23 (1.62)	<0.0001
Consumer price index (monthly)		93.58 (0.58)	93.6 (0.56)	93.35 (0.68)	<0.0001
Consumer confidence index (monthly)		-40.5 (4.63)	-40.59 (4.39)	-39.79 (6.14)	<0.0001
Euribor 3 month rate (daily)		3.62 (1.73)	3.81 (1.64)	2.12 (1.74)	<0.0001
No. of employees (quarterly)		5167.04 (72.25)	5176.17 (64.57)	5095.12 (87.57)	<0.0001

Table 5: Variable Description

Variable	Type	Description
y	Categorical	Client's subscription status to a bank's term Deposit (Yes, No)
age	Quantitative	Age of the Client
job	Categorical	Type of Job (Administration, Blue-collar,Entrepreneur, Housemaid, Management, Retired, Self-Employed, Services, Student, Technician, Unemployed, Unknown)
marital	Categorical	Marital Status (Divorced, Married, Single, Unknown)
education	Categorical	Client's level of education (Basic - 4 years, Basic - 6 years, Basic - 9 years, High School, Illiterate, Professional Coursework, University Degree)
housing	Categorical	Housing Loan Status (Yes, No, Unknown)
Loan	Categorical	Personal Loan Status (Yes, No, Unknown)
contact	Categorical	Contact Communication Type (Cellular, Telephone)
month	Categorical	Last Contact Month of the Year
Day of the Week	Categorical	Last Contact Day of the Week
campaign	Quantitative	Number of Contacts Performed during the Campaign and for the Client
pdays	Quantitative	Number of days that passed by after the client was last contacted from a previous campaign
previous	Quantitative	Number of contacts performed before this campaign and for this client
poutcome	Categorical	Outcome of the previous marketing campaign (failure, non-existent, success)
emp.var.rate	Quantitative	Employment Variation Rate
cons.price.idx	Quantitative	Consumer Price Index
cons.conf.idx	Quantitative	Consumer Confidence Index
euribor 3 m	Quantitative	Daily three month Euribor rate (based on the averaged interest rates at which Eurozone banks offer)
nremployed	Quantitative	Quarterly Average of the Total Number of Employed Citizens