Banking Insurance Product – Phase 3

Blue 18

Ryan Heggie
Ishi Gupta
Lucy Liu
August Majtenyi
Aaryan Sharma

November 26, 2024

Table of Contents

Banking Insurance Product – Phase 3

Overview

The Commercial Banking Corporation (the “Bank”) sought proposals to predict customers' likelihood of purchasing a variable rate annuity, hereafter called the product. Our team previously predicted a customer's likelihood of buying the product with generalized additive models (GAMs), boosting and bagging models, such as random forests (RF) and XGBoosts, and now an explainable boosting model (EBM).

Our team tuned the EBM, evaluated variable importance, and assessed the model's goodness-of-fit metrics via receiver operating characteristic (ROC) curves. We tested the model's performance on the validation set, obtaining an accuracy statistic of 74.1%. Our team decided that the EBM predicts purchase likelihood the best and should be used moving forward.

Key recommendations include marketing the product to those with high balances (over $10,000) in their savings and checking accounts for maximum conversions. Additionally, brand new customers (those with the bank for less than 2 years) and loyal customers (those with the bank for over 34 years) should be targeted, as they have a higher purchasing probability.

Methodology

The following section describes the data used for analysis and how the machine-learning models were created.

Data Used

The Bank provided data about customers who were offered the product. The Bank included a training data set with 8,495 observations and 37 predictor variables relating to the customer's account. We checked for missing values. We used median imputation to fill null values for continuous variables and mode imputation for categorical variables.

EBM

Our team created an EBM to predict the probability of a customer purchasing the product. To reduce overfitting, we incorporated a random variable during feature selection. However, all variables ended up being used in the final EBM. The model was configured to capture up to two-way feature interactions, balancing predictive performance with interpretability. Additionally, we analyzed the importance of global features to find which variables were the most impactful in terms of the EBM. We also found the accumulated local effects (ALE) of account age to highlight how this variable contributed to product purchase predictions.

Analysis

The following section showcases model formation and performance.

EBM Model

We evaluated the EBM on our validation data and created an ROC curve. The ROC curve was used to identify how well the model performed with different threshold values.

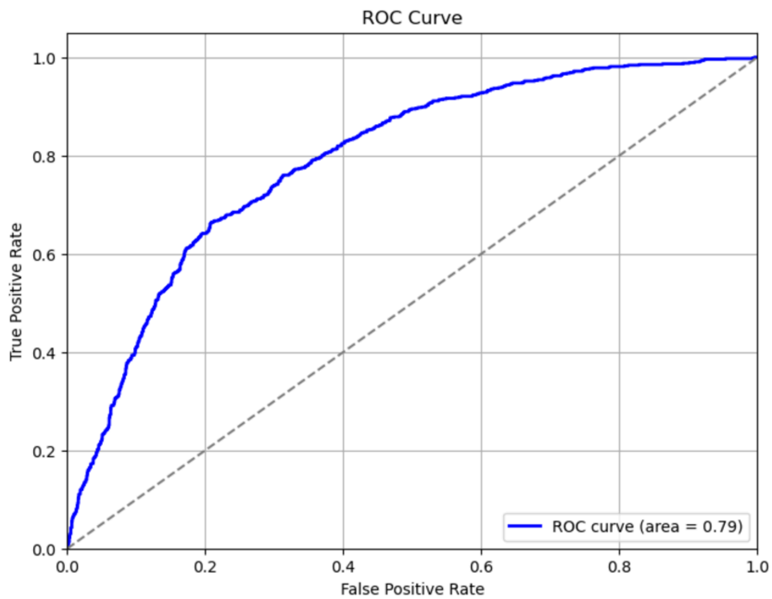


Figure 1: ROC curve of the EBM

The curve above along the diagonal line shows that the EBM's classification performance outperformed random guessing. The area under the curve (AUC) achieved was 0.791, with an optimal cut-off of 0.406.

Final Model Decision

Our final model was chosen based on accuracy measures. We compared EBM with our previous models, as seen in Table 1.

Table 1: Model performance on training and validation data

Rank	Model Name	Accuracy of Validation Data
1	EBM	74.1%
2	RF	72.2%
3	MARS	71.2%
4	GAM	71.1%
5	XGBoost	69.6%
6	Logistic Regression	68.9%

Of the models listed in Table 1, the EBM had an accuracy statistic of 74.1%, successfully predicting product purchasers 74% of the time. Therefore, we chose it as our final model. Based on our final model, we looked at the variable importance.

Global Interpretation of the Final Model

We calculated the global interpretations from our model to find variable importance. We found the global importance score for each variable, as seen in Figure 2.

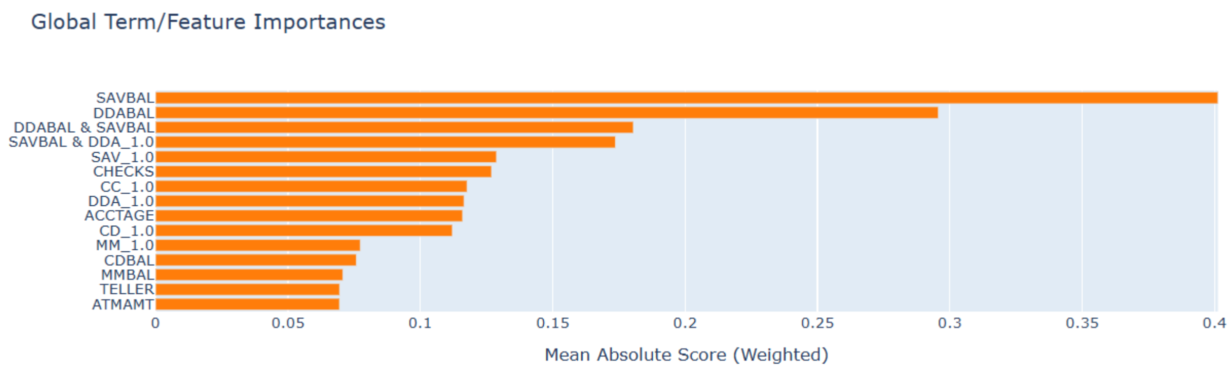


Figure 2: Global variable importance of top variables

According to Figure 2, savings account balance had the largest weighted Mean Absolute Score. This means that, on average, the saving account balance variable moved the natural log of the odds in some direction by 0.4. This average movement was larger than any other variable in terms of impact.

The three most important variables in this model are savings account balance, checking account balance, and the interaction between these two. We recognized that the model heavily emphasized having a certain amount of wealth needed to buy this product. We investigated these variables further with partial dependence plots (PDP).

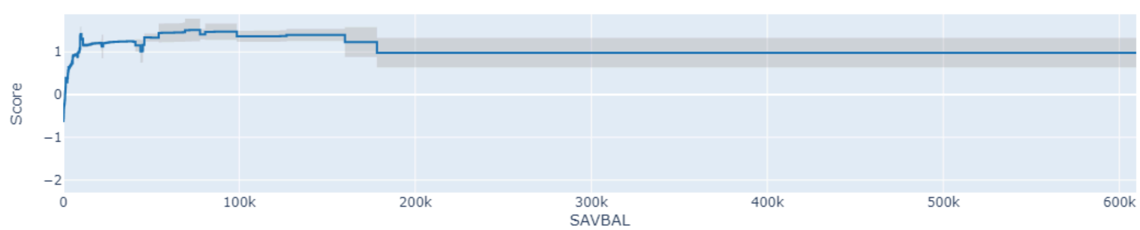


Figure 3: PDP of savings account balance

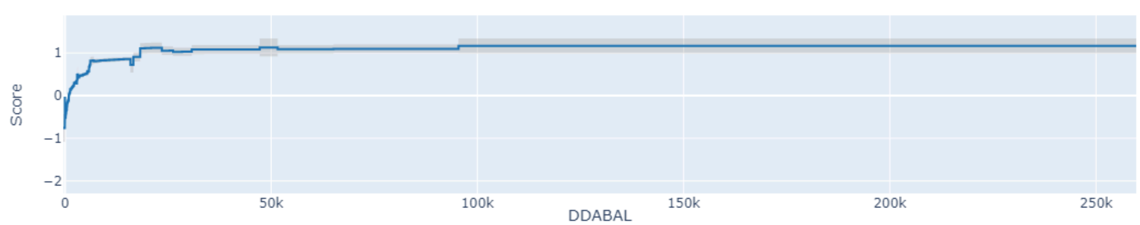


Figure 4: PDP of checking account balance

Figure 3 and Figure 4 show the average score of these variables at different account balance values. Based on the plots, customers with at least $10,000 in either their saving account or checking account are much more likely to purchase the product.

Account Age Interpretation

The ALE analysis for account age revealed distinct customer behavior patterns influencing the likelihood of buying a product. As shown in the Figure 5 below.

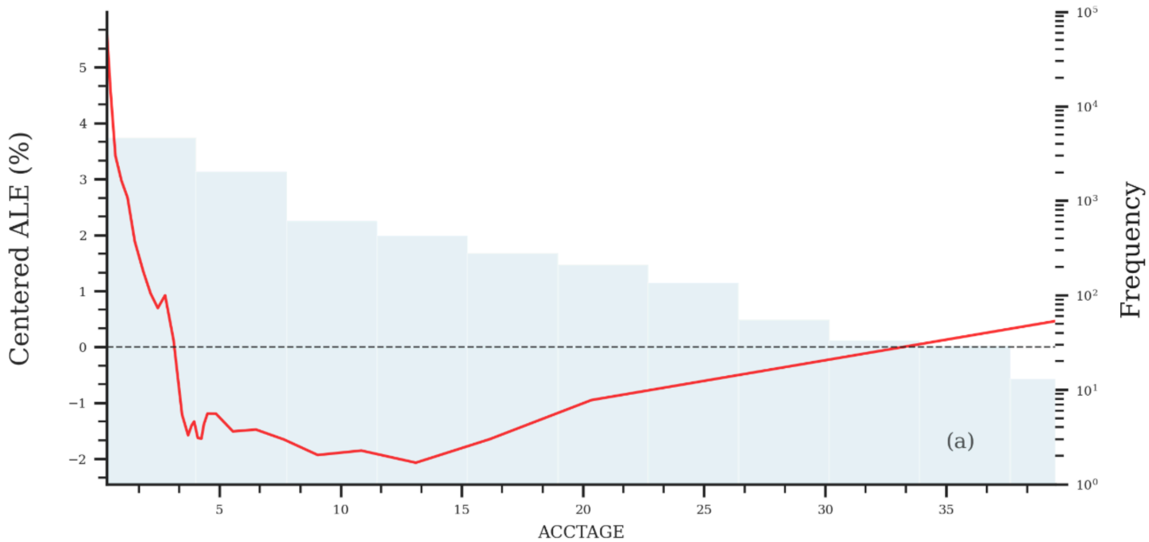


Figure 5: ALE plot for Account Age

As seen in Figure 5, when account age increased from 0 to 14 years, there was a decreasing rate in the probability of purchase, with the sharpest decline between zero and four years. After 14 years, a customer's likelihood of purchasing the product gradually increased. Once the account age reached 34 years, this probability surpassed the average. This indicated that long-standing customers are more likely to purchase the product.

Results & Recommendations

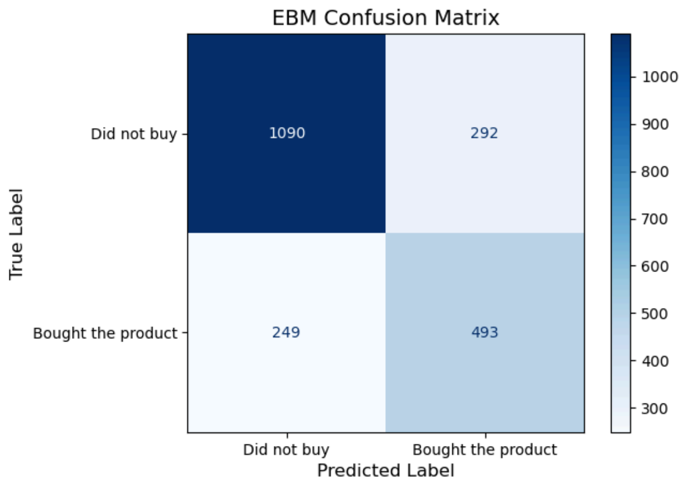
Our EBM was the most accurate model, with an accuracy of 74.1%. Since the most influential variables in this model were savings account balance and checking account balance, we recommend the Bank market to individuals with account balances larger than $10,000. We suggest implementing targeted strategies, enhancing trust-building initiatives, and personalizing engagement for customers with accounts ages below 15 years. The bank should focus marketing efforts and retention strategies on customers with older accounts, particularly those nearing or exceeding 34 years, to maximize conversions.

Conclusion

In this report, we compared our EBM with our four previous models: two GAMs, an RF, and an XGBoost. Since the EBM successfully predicted product purchasers with fewer mistakes, we recommend the Bank use the EBM to make future purchasing predictions.

Our EBM successfully predicted product purchasers 74.1% of the time. Our team identified significant variables – such as savings account balance, checking account balance, and a combined asset amount between the two. Customers with high balances in these accounts should be further investigated. Analyzing the wealth and age of accounts is crucial in determining the likelihood of purchasing the product.

Appendix



Appendix Figure 2: Confusion matrix for the EBM. Sensitivity of 0.664.