Banking on Data

Personal Loan Campaign

Machine Learning

4/19/24

Contents / Agenda

Executive Summary

Actionable Insights for Personal Loan Uptake:

Target High-Income Customers

- Focus on customers with high income.
- Prioritize those with moderate to high credit card expenditures.

Advanced Education Levels

- Direct marketing to customers with higher education degrees.
- Highlight potential financial benefits and loan offers.

Executive Summary (cont)

Great Learning
POWER AHEAD

Strategic Recommendations

Middle-Aged Customer Segment

- Tailor marketing to appeal to middle-aged demographics.
- Emphasize loan benefits relevant to their life stage.

Cross-Selling to CD Account Holders

- Identify customers with CD accounts for loan offers.
- Leverage existing financial relationships for loan promotion.

Leverage Online Banking Data

- Utilize online banking interactions to enhance customer profiles.
- Note that online banking usage is a supplementary, not primary, indicator for loan interest.

ZIP Code Analysis for Market Targeting

- Concentrate on regions within ZIP code categories 94 and 92.
- These areas show a higher rate of personal loan conversion.

Business Problem Overview and Solution Approach

Problem Statement & Solution Methodology

The Challenge at AllLife Bank

- Previous campaigns yielded a 9% conversion rate for personal loans among liability customers.
- Objective: Enhance this success by pinpointing potential loan customers within our current base.

Our Analytical Approach

- Deployed a Decision Tree Classifier to dissect and understand customer data.
- Iteratively refined model performance through hyperparameter optimization.
- Gained insights into key customer attributes that drive loan acquisition.

Insights and Key Predictors

- Identified top predictors for personal loan uptake: income level, family size, and education.
- Feature importance analysis underpinned our targeted approach, optimizing our future campaigns.

EDA Results

EDA Key Findings & Age Analysis for Loan Uptake

Significant Predictors Identified

- Income, credit card spending (CCAvg), and higher education levels stand out as top indicators for loan purchases.

Target Age Groups for Effective Marketing

- Interest in loans spikes among customers in their early to mid-30s and mid-40s.
- Age distributions suggest these are optimal groups for targeted campaigns.

Non-Loan vs. Loan Customer Age Trends

- Non-loan customers: Age is bell-shaped with a peak in the 40s, indicating a widespread age range.
- Loan customers: Similar bell-shape but slightly skewed towards younger ages, implying middle-aged individuals have a higher tendency for loan uptake.

EDA Results (cont)

Age Distribution & Education Level Insights

Non-Loan Customer Age Profile:

- Median age in the late 30s to early 40s.
- Wide age distribution implies a diverse customer base.

Loan Customer Age Characteristics:

- Median age similar to non-loan customers.
- Narrower interquartile range points to a specific age group with higher loan uptake.
- Education's Impact on Loan Acquisition

Correlation with Education:

- Higher education aligns with increased loan acquisition.
- Greater financial literacy and earning potential among highly educated individuals.

Marketing Strategy Recommendations:

- Target customers with higher education for loan products.
- Emphasize financial solutions suited to life stages of the educated demographic.
- Utilize education as a key segment criterion in personalized marketing efforts.

Data Preprocessing

<u>Data Quality Assurance</u>

Duplicate Value Check:

- Utilized data.duplicated() to ensure no duplicates exist, confirming the dataset's quality for modeling.

Data Anomalies & Treatment-

Missing Value Handling:

- Utilized data.isna().sum() to ensure no missing values exist, confirming the dataset's quality for modeling.

Outlier Management:

- Replaced the negative numbers/outliers in 'Experience’ with positive values.

Data Preprocessing (cont)

Data Quality Assurance

Refined Feature Engineering-

ZIP Code Grouping:

- Consolidated into regional clusters for sharper geographic analysis.

Data Preprocessing (cont)

<u>Data Preprocessing for Model Readiness</u>

Categorical Variable Transformation:

- Applied OneHotEncoding to 'Education' and 'Family' to convert them into a model-friendly binary format.

Preparation for Decision Tree Modeling:

- These preprocessing steps were critical in developing a decision tree model primed to forecast personal loan uptake effectively.

These actions ensured the dataset was optimally prepared for the modeling phase, setting the stage for accurate prediction results.

Model Building

Building the Decision Tree Model

Data Preparation:

- Preprocessed data and handpicked features crucial for loan prediction.

Model Training:

- Developed the model using the selected features and trained it on the dataset.

Hyperparameter Tuning:

- Optimized performance by fine-tuning the model's hyperparameters.

Model Building (cont)

<u>**Pruning and Validating the Model**</u>

Pruning:

- Implemented to streamline the model, targeting a balance in complexity and preventing overfitting.

Validation:

- Employed cross-validation to ensure robustness and generalization of the model.
- Performance Evaluation: Assessed using accuracy and other key metrics to ensure precision.

Performance Evaluation:

- Assessed using accuracy and other key metrics to ensure precision.

Model Building (cont)

Model Performance Outcomes

On Training Data:

- Achieved 90.54% true negatives and 9.46% true positives.
- Zero false classifications, indicating an ideal accuracy in the training set.

On Testing Data:

- Accurate 'no loan' predictions in 1342 cases, and 'loan' in 133 cases.
- Noted 16 false negatives and 9 false positives, a sign of minor errors.
- Maintained a high accuracy rate, reflected in a robust F1 score.

Model Performance Summary

Decision Tree Model Overview

Evaluation Metrics: Accuracy, precision, F1 score, and recall were used to gauge the model's predictive performance across both training and testing sets.

Overall Correctness: These metrics provided a comprehensive view of the model's correctness in classification.

Income as Key Predictor: With an income threshold of 98.5, the model prioritizes income level in its predictions.

Gini Impurity Analysis: The significant drop in Gini impurity from the root to the left node indicates a well-defined customer segment, while the right node's higher impurity points to a more diverse group.

Conclusion: The decision tree's reliance on income as a primary factor ensures focused predictions on loan purchases, with the model's calibration on impurity levels fine-tuning its sensitivity and precision.

Model Performance Summary (cont)

Key Predictive Features: Income, family size, education level, and CCAvg surfaced as essential in forecasting loan purchases.

Tabular Comparison: Below is a table summarizing accuracy, precision, F1 score, and recall across all models for both training and test data.

Performance Insights: Enable a clear comparison to identify the best-performing model.

Training performance comparison:

	Decision Tree sklearn	Decision Tree (Pre-Pruning)
Accuracy	1.0	0.987714
Recall	1.0	0.873112
Precision	1.0	0.996552
F1	1.0	0.930757

Test performance comparison:

	Decision Tree sklearn	Decision Tree (Pre-Pruning)
Accuracy	0.983333	0.978667
Recall	0.892617	0.785235
Precision	0.936620	1.000000
F1	0.914089	0.879699

Model Performance Improvement

Enhancing Model Performance through Pruning

- **Pruning Techniques:** Implemented various pruning strategies to optimize the decision tree.
- **Complexity Adjustment:** Adjusted the alpha parameter to reduce overfitting, boosting the model's generalizability to new data.
- **Refined Focus:** Pruning concentrated the model on vital predictors, removing less informative branches, thus enhancing efficiency and effectiveness.

Decision Rules and Feature Importance

- **Rule Application:** Utilized specific decision rules determined by the feature importance to guide data splits and predictions.
- **Importance Analysis:** Feature importance analysis was crucial in identifying key predictors such as income and education, informing the pruning strategy.
- **Predictive Accuracy:** Focused adjustments to the model based on these insights significantly improved predictive accuracy, making the model adept at handling real-world data.

APPENDIX

Appendix

Data Preprocessing Details:

- A systematic approach was used to handle missing values, with negative values in the 'Experience' column and replaced with positive values.
- Subcategorized the ZIPCode column into groups using their first two number.
- Outliers in variables like 'Income', 'CCAvg', and 'Mortgage' were retained without transformation to represent high-income customers.

Model Building and Evaluation Details:

- The Decision Tree Classifier model was chosen, optimized using GridSearchCV for hyperparameters like 'max_depth', 'min_samples_leaf', and 'max_leaf_nodes'.
- Model performance was assessed on both training and testing data, with a focus on accuracy, recall, precision, and F1 score.
- Pruning was used to refine the model, selecting the best ccp_alpha value to balance complexity and prevent overfitting.

Appendix (cont)

Data Preprocessing Details (cont)

Feature Importance Analysis:

- Features such as 'Income', 'Family', 'Education', and 'CCAvg' were identified as the most significant in predicting personal loan uptake.

Model Performance Data:

- The confusion matrix indicated a high true negative rate and a moderate true positive rate, with very few false positives and negatives, suggesting a conservative model that prioritizes accuracy in predicting non-purchasers.

Demographic Insights:

- Analyses of various customer attributes, including 'Age', 'Family', 'Experience', 'Income', and 'Education', provided insights into their influence on the probability of purchasing a loan.

Data Background and Contents

<u>**Customer Profile:**</u> The dataset includes comprehensive demographic and financial data from 5,000 customers.

Key Attributes:

Demographic Details: Age, Experience, Family Size, ZIP Code.

Financial Metrics: Income, CCAvg (monthly credit card expenditure), Mortgage.

Banking Services: Personal Loan status, Securities Account, CD Account, Online banking usage, and Credit Card usage.

Purpose: To analyze customer behavior and predict personal loan uptake within the existing customer base of AllLife Bank.

Additional Recomendations

To increase personal loan customers across all three different income groups, the following strategies could be considered:

For Lower Income Group:

- Deploy educational campaigns on personal loan benefits to raise awareness.
- Offer small, easy-to-repay loans tailored for low-income individuals.
- Highlight the usability of loans for essential personal developments, like education or home improvement.

For Mid-Income Group:

- Focus on the flexibility and competitive interest rates of loans to appeal to middle-class aspirations.
- Introduce bundled offers, like personal loans with discounted insurance or investment products.
- Create referral programs encouraging existing customers to refer peers, potentially within the same income bracket.

For Higher Income Group:

- Highlight the convenience and potential tax benefits associated with personal loans.
- Offer premium services, like dedicated account managers or expedited loan approval processes.
- Leverage existing data to personalize marketing efforts, appealing to their lifestyle & financial goals.

Great Learning

Happy Learning !