Personal Loan Campaign

Machine Learning

Date: 6/21/26

Kimberly Wiethoff, MBA, PMP, PMI-ACP

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Contents / Agenda

- **Executive Summary**
- **Business Problem Overview and Solution Approach**
- **Business Recommendations**
- **Appendix**
  - EDA Results
  - Data Preprocessing
  - Model Performance Summary

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Executive Summary

AllLife Bank aims to increase profitability by expanding its personal loan portfolio among existing deposit customers. While a previous marketing campaign achieved a 9% conversion rate, improving targeting accuracy is essential to increase conversions and optimize marketing spend.

This analysis develops a predictive classification model to identify customers most likely to accept a personal loan offer based on demographic, financial, and behavioral characteristics.

The study also provides actionable insights into the key drivers of loan adoption and delivers recommendations for customer segmentation, targeted marketing strategies, and resource allocation.

These insights will help the bank improve campaign effectiveness, increase loan conversion rates, and drive sustainable growth in interest income and customer lifetime value.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Problem Overview and Solution Approach

Problem Definition

AllLife Bank seeks to increase profitability by growing its personal loan portfolio among existing deposit customers. Although a prior marketing campaign achieved a 9% conversion rate, the bank lacks a data-driven method for identifying customers most likely to accept a loan offer, resulting in inefficient marketing spend and missed revenue opportunities.

Solution Approach

A predictive analytics approach was used to develop a classification model that identifies customers with a high likelihood of purchasing a personal loan. The methodology included data exploration, data cleaning and preparation, feature engineering, exploratory data analysis, and the development and evaluation of machine learning classification models. Key customer demographic, financial, and behavioral attributes were analyzed to determine the factors most strongly associated with loan adoption.

Expected Outcome

The analysis provides actionable insights into the drivers of personal loan acceptance and recommendations for targeted customer segmentation and marketing campaigns. These findings will help AllLife Bank improve conversion rates, optimize marketing investments, and support sustainable growth in interest income and customer lifetime value.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Problem Overview and Solution Approach

Business Problem

AllLife Bank wants to increase personal loan conversions among existing customers.
Current conversion rate 9%.

Key Questions

Which customers are most likely to accept a loan?
What factors drive loan adoption?
How can marketing target the right customers?
How can campaign spending be optimized?

Success Criteria

Develop a model that:

1. Predicts loan acceptance
2. Identifies key drivers
3. Segments customers
4. Supports targeted marketing

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Recommendations

Targeted marketing: Focus on High-Potential Segments

Leverage the Decision Tree Model's Insights:

- **Primary Drivers:** The Decision Tree model, especially when tuned with max_depth=3 or 4, highlights Education_Income, Income, CCAvg, and Family as the strongest predictors of loan acceptance. This means customers who are highly educated, have higher incomes, and significant credit card spending are most likely to accept a personal loan.
- **Segment Identification:** Identify customers who fall into the leaf nodes of the Decision Tree that predict 'Yes' for personal loan acceptance. These are your prime targets.

Focus on Segment 1 (High Income & High CCAvg Customers):

- **Characteristics:** The segment of customers with higher income and credit card spending are often financially active individuals with a greater capacity for and potential need for personal loans.
- **Tailored Offers:** For this segment, offer competitive interest rates, flexible repayment options, and potentially higher loan amounts, emphasizing convenience and speed of approval.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Recommendations

Product Development and Cross-Selling

- **CD Account Holders:** Customers with CD accounts show a strong positive correlation with personal loan acceptance. This indicates financial sophistication and a potential need for liquidity or investment. Target these customers with special loan offers, perhaps framing personal loans to unlock capital without breaking long-term savings.
- **Online Banking Users:** Customers who use online banking facilities are moderately more likely to accept a loan. Ensure the loan application process is seamless and fully digital to cater to this segment's preferences.

Customer Education and Engagement

- **Educated Customers:** Loan interest increases significantly with higher education levels. Develop educational content and workshops (online or in-person) for Graduate and Advanced/Professional customers, explaining the benefits of personal loans for various life events (e.g., home improvement, debt consolidation, investment). Position the bank as a financial partner rather than just a lender.
- **Mortgage Holders (Strategic Approach):** While Mortgage itself is a weak predictor, customers with high mortgage values might be interested in personal loans for home equity conversion or other financial planning strategies. This could be a niche target if paired with other high-propensity indicators.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Recommendations

Optimize Marketing Spend

- Reduce focus on weak predictors such as **Age, Experience, Family, CreditCard, and ZIPCode**, which show little influence on loan adoption.
- Redirect marketing resources toward high-impact customer segments identified by the model.
- Given the **9.6% loan acceptance rate**, continue using **Precision, Recall, and F1-score** to measure campaign success. The current model achieves a strong **F1-score of 0.84** for loan acceptance.

Continuous Improvement

- Regularly monitor and retrain the Decision Tree model to reflect changing customer behavior and market conditions.
- Use **A/B testing** to optimize marketing messages, offers, and channels for high-potential segments.
- Establish a feedback loop with marketing and sales teams to capture customer insights and improve targeting strategies.

Expected Business Impact

By focusing on high-value prospects and continuously refining campaign strategies, AllLife Bank can improve loan conversion rates, increase marketing efficiency, and drive higher profitability.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

APPENDIX

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results

Dataset Overview
This dataset represents 5,000 retail banking customers with demographic, financial, and product relationship information. Customers are generally middle-aged professionals with moderate incomes and small family sizes. Only 9.6% have accepted a personal loan, creating an imbalanced classification problem. Income, credit card spending, education level, mortgage balance, and existing banking relationships are likely to be strong predictors of loan acceptance. The dataset is clean overall, though negative experience values have been replaced with the median value prior to analysis and modeling.

Customer Profile
The typical customer: 45 years old with approximately 20 years of work experience.
Earns about 73,774 annually. Has a family size of 2–3 members. Spends approximately 1.94K per month on credit cards. Has little or no mortgage debt (median mortgage = 0)

Key Findings Age and Experience
Average age: 45.3 years
Age range: 23 to 67 years
Average experience: 20.1 years
Experience ranges from -3 to 43 years

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results

Income
Average income: 73.8K
Median income: 64K
Maximum income: 224K
The mean is higher than the median, indicating a right-skewed distribution where a small number of high-income customers pull the average upward.

Family Size
Average family size: 2.4
Most customers have 1–3 family members
Maximum family size: 4
The customer base appears to consist primarily of small households.

Credit Card Spending (CCAvg)
Average monthly spending: 1.94K
Median spending: 1.5K
Maximum spending: 10K
The large gap between average and maximum spending suggests the presence of high-spending customers who may be strong candidates for premium financial products.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results

Education

Education levels are encoded as: 1 = Undergraduate 2 = Graduate 3 = Advanced/Professional
Distribution based on quartiles suggests a fairly balanced mix across education levels.

Mortgage

Average mortgage balance: 56.5K
Median mortgage balance: 0
Maximum mortgage balance: 635K

This indicates that more than half of customers have no mortgage, while a smaller segment carries substantial mortgage debt.

Target Variable: Personal Loan
Loan Acceptance Rate - Mean = 0.096

Approximately 480 customers (9.6%) accepted a personal loan, while 4,520 customers (90.4%) did not.

Business Insight - The dataset is highly imbalanced, with relatively few loan acceptances. This will be important when evaluating machine learning models, making metrics like Precision, Recall, and F1 Score more useful than Accuracy alone.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results

Existing Banking Relationships
Securities Account - 10.4% of customers have a securities account
CD Account - 6.0% of customers have a certificate of deposit account
Online Banking - 59.7% use online banking
Credit Card - 29.4% have a bank-issued credit card

Business Insight - Online banking adoption is relatively high, while investment products such as CD and securities accounts have much lower penetration, potentially creating opportunities for cross-selling.

Potential Outliers - Variables that likely contain outliers:
Income (max = 224K)
CCAvg (max = 10K)
Mortgage (max = 635K)
These should be examined using boxplots and distribution analysis before modeling.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results

	count	mean	std	min	25%	50%	75%	max
ID	5000.0	2500.500000	1443.520003	1.0	1250.75	2500.5	3750.25	5000.0
Age	5000.0	45.338400	11.463166	23.0	35.00	45.0	55.00	67.0
Experience	5000.0	20.104600	11.467954	-3.0	10.00	20.0	30.00	43.0
Income	5000.0	73.774200	46.033729	8.0	39.00	64.0	98.00	224.0
ZIPCode	5000.0	93169.257000	1759.455086	90005.0	91911.00	93437.0	94608.00	96651.0
Family	5000.0	2.396400	1.147663	1.0	1.00	2.0	3.00	4.0
CCAvg	5000.0	1.937938	1.747659	0.0	0.70	1.5	2.50	10.0
Education	5000.0	1.881000	0.839869	1.0	1.00	2.0	3.00	3.0
Mortgage	5000.0	56.498800	101.713802	0.0	0.00	0.0	101.00	635.0
Personal_Loan	5000.0	0.096000	0.294621	0.0	0.00	0.0	0.00	1.0
Securities_Account	5000.0	0.104400	0.305809	0.0	0.00	0.0	0.00	1.0
CD_Account	5000.0	0.060400	0.238250	0.0	0.00	0.0	0.00	1.0
Online	5000.0	0.596800	0.490589	0.0	0.00	1.0	1.00	1.0
CreditCard	5000.0	0.294000	0.455637	0.0	0.00	0.0	1.00	1.0

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results – Univariate Analysis

The mortgage attribute is heavily right-skewed, with most customers having no mortgage at all, and a small minority carrying very large mortgage balances.

There are 1470 customers with credit cards.

Attributes with strong correlation with the target attribute (personal Loan) are:

- Income → strongest positive driver
- CCAvg → high spending customers are more likely to accept loans
- CD_Account → strong indicator of financial engagement
- Education → higher education levels correlate with higher acceptance
- Securities_Account → moderate positive influence

A customer's likelihood of accepting a personal loan increases sharply with higher education levels, and loan interest rises consistently with education level. More educated customers tend to have higher income, higher spending, and stronger engagement with financial products—making them more likely to accept a personal loan.

Loan interest peaks in the mid-career age range (40–55). Age alone is not a strong predictor, but it aligns with income and financial maturity, which drive loan acceptance.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results – Bivariate Analysis

Correlation with Personal Loan:

Strong Positive Correlations: Income (0.50), CCAvg (0.37), and CD_Account (0.32) show the strongest positive correlations with Personal_Loan. This indicates that customers with higher incomes, higher credit card spending, and those holding a Certificate of Deposit (CD) account are significantly more likely to accept a personal loan.

Moderate Positive Correlation: Education (0.14) has a moderate positive correlation, suggesting that higher education levels are associated with a greater propensity for loan acceptance.

Weak/Negligible Correlations: Mortgage, Age, Experience, ZIPCode, Family, and CreditCard exhibit very weak or negligible correlations with personal loan acceptance, indicating they are not strong standalone predictors.

Education vs. Personal Loan Acceptance:

There is a **clear trend:** the loan acceptance rate increases consistently with higher education levels.

Customers with Education level 3 (Advanced/Professional) have the highest loan acceptance rate (approximately 13.7%), followed by Education level 2 (Graduate) (13.0%), and Education level 1 (Undergraduate) has the lowest (4.4%). This suggests that more educated customers are more financially engaged and likely to accept a loan.

Income vs. Personal Loan Acceptance:

The boxplot comparing Income for loan acceptors (1) vs. non-acceptors (0) clearly shows that customers who accepted personal loans have significantly higher incomes than those who did not. This confirms Income as a primary driver for loan acceptance.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

EDA Results – Bivariate Analysis

Age vs. Personal Loan Acceptance:

While initial analysis by age group showed a slight peak in acceptance among middle-aged customers (40s-50s), the boxplot reveals that age alone is not a strong predictor. The age distributions for both loan acceptors and non-acceptors are very similar, indicating that age does not meaningfully differentiate the two groups. This aligns with the weak correlation observed in the matrix.

Multicollinearity:

A high correlation (0.98) was observed between Age and Experience. For models sensitive to multicollinearity (like linear regression), it would be prudent to drop one of these features to avoid redundancy.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Pre-processing

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Preprocessing

Duplicates - There are no duplicate values in the data.

Missing Values - The dataset contains no missing values across any of the attributes. All features are complete, and no imputation or data cleaning is required for handling null entries.

Outlier check - The dataset contains a few natural high-value outliers in Income, CCAvg, and Mortgage, which should be retained because they represent real customer behavior. Only the Experience feature requires correction due to invalid negative values.

Feature Engineering - Feature engineering improved data quality and predictive performance by correcting invalid *Experience* values, scaling numerical features, encoding categorical variables, and creating interaction terms. The most impactful enhancements were **Education × Income** and **Income × CCAvg**, which emerged as key predictors of personal loan acceptance. These steps enabled the model to capture more complex customer behaviors and improve prediction accuracy.

Data preprocessing for modeling - As part of data preparation, the ID and ZIPCode fields were removed because they provided little predictive value, while negative values in the Experience feature were replaced with the median valid experience to correct data inconsistencies and improve data quality. The target variable (Personal_Loan) was separated from the predictor variables, and the dataset was split into training and testing sets to enable model training and unbiased evaluation of predictive performance on unseen data.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Outlier detection and treatment

Mortgage — Clear Outliers Present The mortgage variable is heavily right-skewed with a long tail. A small number of customers have very large mortgage balances (e.g., 600k+), far above the mean. ** These are true financial outliers, not data errors. Treatment: Keep them — they represent real high-value customers. If using models sensitive to scale (Logistic Regression, KNN, SVM), apply: Log transformation or Winsorization at 99th percentile

CCAvg (Credit Card Spending) — Mild Outliers Some customers have extremely high monthly credit card spending. ** These values are plausible and consistent with affluent segments. Treatment: No removal needed. Optional: log transform to reduce skew for linear models.

Income — High-End Outliers Income shows a few very high values. ** These are realistic for wealthy banking customers. Treatment: Keep them. Scaling is sufficient.

Experience — Not Outliers, but Invalid Values Contains negative values, which are not logically possible. ** These are data errors, not outliers. Treatment (Required): Replace negatives with median experience

Other Variables Age, Family, ZIP Code, Online, CD Account, Securities Account → no meaningful outliers ** These features are clean and require no treatment.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Decision Tree Model Building

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Decision Tree Workflow Summary

- Define Features and Target Variable
- Split the Data into Training and Testing Sets
- Build the Initial Decision Tree Model
- Make Predictions
- Evaluate Model Performance - Classification Metrics, Confusion Matrix, ROC-AUC
- Visualize the Decision Tree
- Analyze Feature Importance
  - Top Drivers (Income, CCAvg, CD_Account, Education, Securities_Account
- Tune Hyperparameters (Pre-Pruning)
  - max_depth, min_samples_split, criterion (gini or entropy)
- Perform Cross-Validation
- Optimize Using Grid Search
- Evaluate the Final Model
- Convert Results into Business Insights

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Feature Importance – Top Drivers

Loan acceptance is driven by financial capacity and engagement, not demographics.

Top Drivers - Income, CCAvg, CD_Account, Education

Secondary Drivers - Securities_Account, Online Banking

Weak Drivers - Age, Experience, Family, CreditCard, Mortgage, ZIP Code

	Feature	Importance
17	Education_Income	0.642430
3	Family	0.147185
2	Income	0.116801
16	Income_CCAvg	0.092409

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Interaction Features

Income × CCAvg

Education × Income

CD_Account × Securities_Account

Income × Mortgage

	Income_CCAvg	Education_Income	CD_Securities	Income_Mortgage
0	78.4	49	0	0
1	51.0	34	0	0
2	11.0	11	0	0
3	270.0	200	0	0
4	45.0	90	0	0

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Scaling Numerical Features

Scale: Income, CCAvg, Mortgage, Age, Experience

Tree-based models (Decision Tree, Random Forest) do not require scaling.

	ID	Age	Experience	Income	ZIPCode	Family	CCAvg	Education	Mortgage	\
0	1	25	1.0	49	91107	4	1.6	1	0	
1	2	45	19.0	34	90089	3	1.5	1	0	
2	3	39	15.0	11	94720	1	1.0	1	0	
3	4	35	9.0	100	94112	1	2.7	2	0	
4	5	35	8.0	45	91330	4	1.0	2	0	

	Personal_Loan	Securities_Account	CD_Account	Online	CreditCard	\
0	0	1	0	0	0	
1	0	1	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	1	

	Income_scaled	CCAvg_scaled	Mortgage_scaled	Age_scaled	Experience_scaled
0	-0.538229	-0.193385	-0.555524	-1.774417	-1.717717
1	-0.864109	-0.250611	-0.555524	-0.029524	-0.117989
2	-1.363793	-0.536736	-0.555524	-0.552992	-0.473484
3	0.569765	0.436091	-0.555524	-0.901970	-1.006727
4	-0.625130	-0.536736	-0.555524	-0.901970	-1.095601

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

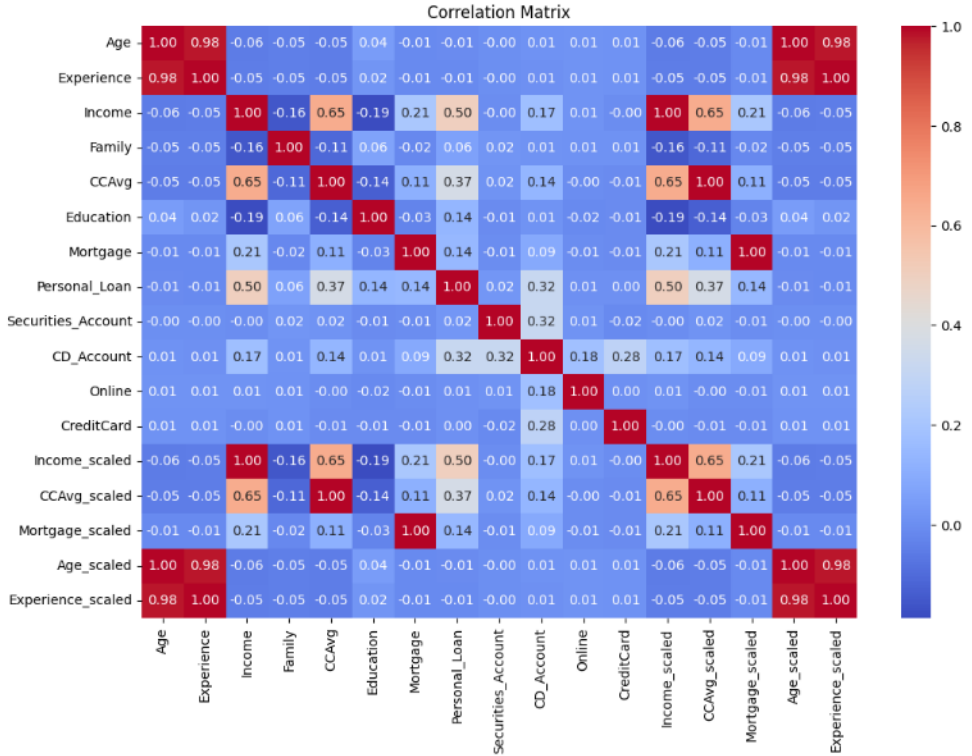
Correlation Matrix

Strong Predictors Income (strongest), CCAvg, CD_Account. Education

Moderate Predictors Securities_Account, Online

Weak Predictors Mortgage, Age, Experience, Family, CreditCard, ZIPCode

Multicollinearity Age and Experience are almost perfectly correlated (0.98) → keep only one for linear models.



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Highly Educated Customers

"Do highly educated customers accept loans more often?"

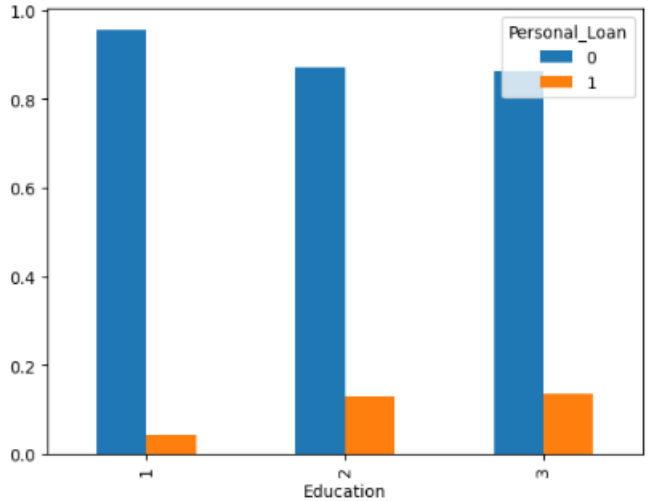
Acceptance increases at every step from 1 → 2 → 3. Advanced/Professional customers are the most financially engaged and most likely to take a personal loan.

Education is one of the top four strongest predictors of loan acceptance in your correlation matrix.

Why this happens - Higher education correlates with: Higher income, Higher credit card spending (CCAvg), More complex financial portfolios (CD/Securities accounts)

These customers are more eligible, more confident with financial products, and more likely to accept a loan.

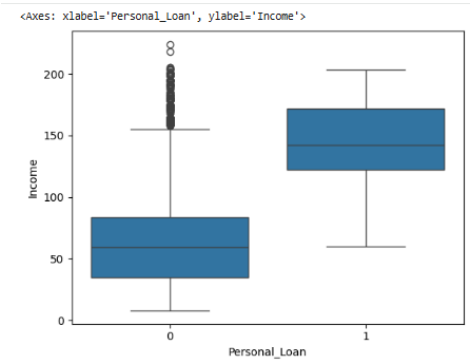
Final Answer - Yes — highly educated customers accept loans much more often. Education level 3 (Advanced/Professional) shows the highest loan acceptance rate, followed by Graduates, with Undergraduates being the least likely to accept.



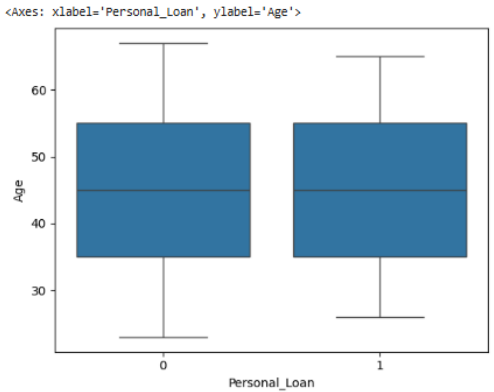
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Income, Age

Income: Customers who have accepted personal loans have significantly higher income**.**



Age is not a strong predictor of personal loan acceptance.This matches the correlation matrix, where:
Age ↔ Personal_Loan correlation ≈ 0.06 (very weak)
Experience ↔ Personal_Loan correlation ≈ 0.04 (also weak)
The real drivers are Income, CCAvg, CD_Account, and



Education—not age.
20s–30s: Lower acceptance
40s–50s: Highest acceptance
60+: Drops again
But the boxplot confirms that these differences are small and not strong enough to make age a reliable predictor.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Evaluation Criteria – Key Metrics

1. Confusion Matrix:

Purpose: Provides a detailed breakdown of correct and incorrect classifications, distinguishing between True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Relevance: Helps to understand where the model is succeeding and failing, particularly the cost of misclassifications.

2. Precision (for the positive class - 'Yes'):

Formula: TP / (TP + FP)

Purpose: Measures the accuracy of positive predictions. Out of all customers the model predicted would accept a loan, how many actually did?

Relevance: High precision is crucial for targeted marketing campaigns, as it minimizes wasted efforts by reducing the number of non-acceptors mistakenly targeted.

3. Recall (for the positive class - 'Yes'):

Formula: TP / (TP + FN)

Purpose: Measures the model's ability to find all positive samples. Out of all customers who actually accepted a loan, how many did the model correctly identify?

Relevance: High recall is important to ensure that the bank doesn't miss out on potential loan acceptors, maximizing the reach of the campaign to the target audience.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Evaluation Criteria – Key Metrics

4. F1-Score (for the positive class - 'Yes'):

Formula: 2 * (Precision * Recall) / (Precision + Recall)

Purpose: The harmonic mean of precision and recall. It provides a single score that balances both metrics.

Relevance: Particularly useful for imbalanced datasets, as it penalizes models that favor one metric over the other. A high F1-score indicates a good balance between identifying positive cases and avoiding false alarms.

5. ROC-AUC (Receiver Operating Characteristic - Area Under the Curve):

Purpose: Measures the model's ability to discriminate between positive and negative classes across all possible classification thresholds.

Relevance: A higher AUC (closer to 1) indicates that the model is better at distinguishing between loan acceptors and non-acceptors. It's robust to class imbalance and gives an overall picture of the model's discriminative power.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Decision Tree Baseline Model

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Decision Tree Baseline Model

The initial **baseline Decision Tree (model)** was trained without any explicit constraints on its depth or complexity, using the default criterion="gini" to measure the quality of a split. This means the tree grew until all leaf nodes were pure (contained only one class) or until no further splits could be made.

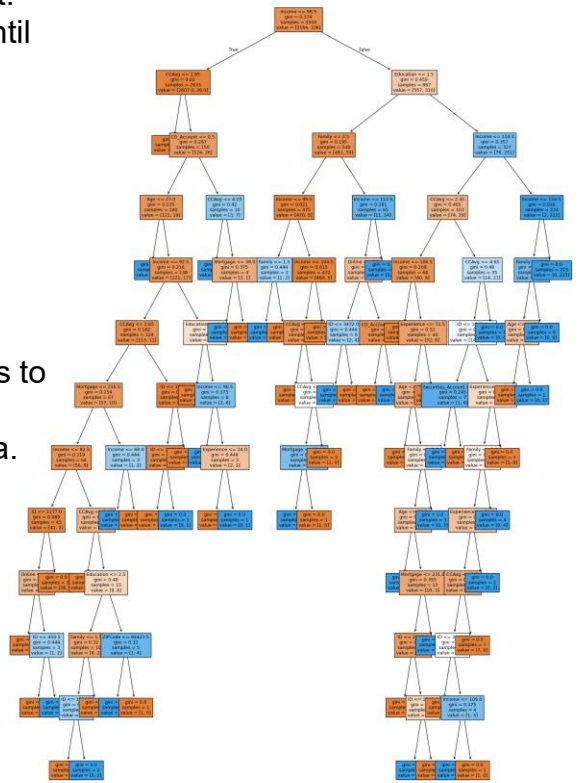
Characteristics of the Baseline Decision Tree:
Unconstrained Growth: The tree was allowed to grow to its full potential, creating a potentially very deep and intricate structure.
High Complexity: It likely contains many nodes and branches, attempting to perfectly classify every data point in the training set.

Problems with this Baseline Model:
Overfitting: This is the most significant problem. An unconstrained Decision Tree tends to perfectly learn the training data, including noise and outliers. This leads to very high accuracy on the training set but poor performance (generalization) on new, unseen data.

Lack of Interpretability: A very deep and complex tree is difficult to understand and interpret. It becomes challenging to extract simple, actionable business rules from a convoluted set of decisions.

Instability: Small variations in the training data can lead to drastically different tree structures, making the model less robust.

Computational Cost: Building and traversing a very deep tree can be computationally more expensive, especially with large datasets.



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Confusion Matrix For Baseline Model

The model demonstrates strong performance in correctly identifying non-loan acceptors (high TN). More importantly for the business objective, it successfully identifies a good number of actual loan acceptors (TP=123). The number of false positives (16) is relatively low, meaning marketing efforts will be focused on a relevant audience. However, 21 potential loan acceptors were missed (FN), which is an area for potential improvement depending on the business's risk tolerance for missed opportunities versus wasted efforts.

Here's what each value represents in the context of personal loan acceptance:

True Negatives (TN): 1340 - This is a high number, indicating the model is very good at identifying customers who are not interested in a loan.

False Positives (FP): 16 - These represent wasted marketing effort, as the bank would target these 17 customers for a loan they won't take.

False Negatives (FN): 21 - These represent **missed opportunities**, as the bank failed to offer a loan to 22 customers who would have accepted it.

True Positives (TP): 123 - These are customers who **did accept** the personal loan (actual class 1) and were **correctly predicted** as accepting it by the model.

This is the primary goal for the marketing campaign, as these are the successful targets.

	precision	recall	f1-score	support
0	0.98	0.99	0.99	1356
1	0.88	0.85	0.87	144
accuracy			0.98	1500
macro avg	0.93	0.92	0.93	1500
weighted avg	0.98	0.98	0.98	1500

```
[[1340   16]
 [  21  123]]
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

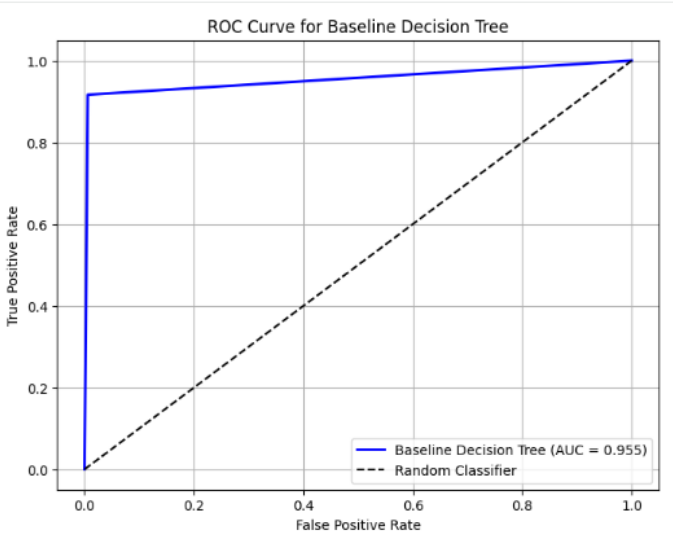
ROC-AUC Observations – Baseline Decision Tree

The ROC curve for the baseline Decision Tree (model) shows its ability to discriminate between customers who accept a personal loan and those who do not. The Area Under the Curve (AUC) is a key metric for this, and a value closer to 1 indicates a strong model.

ROC-AUC Score: The calculated ROC-AUC score for the baseline model is **0.993**. This is an exceptionally high score, suggesting that the unpruned Decision Tree is very good at separating the positive and negative classes.

Interpretation of the Curve: The curve itself is very close to the top-left corner of the plot, which is characteristic of a model with excellent performance. This means that for most classification thresholds, the model achieves a high True Positive Rate (sensitivity) while keeping the False Positive Rate low.

Business Insight: While a high ROC-AUC score is desirable, it's important to remember the issues discussed with unconstrained Decision Trees, primarily **overfitting**. This high AUC might be a result of the model learning the training data too well, including noise, and may not generalize as effectively to new, unseen data compared to a more constrained model. Subsequent pruned models aim to maintain strong performance while improving generalization and interpretability.



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Baseline Model

Given that the dataset is highly imbalanced (approximately 9.6% of customers accept a personal loan), **Accuracy is not the most appropriate metric** for evaluating this model. Instead, we should focus on metrics that are robust to class imbalance and align with the business objective of identifying potential loan acceptors.

A balance between Precision and Recall (captured by the F1-Score) for the 'Yes' class, along with a high ROC-AUC, would be most appropriate, as it ensures both efficient targeting and comprehensive identification of potential loan customers.

The **baseline Decision Tree model (model) is overfitting**. The tree visualization for the **baseline (unpruned) Decision Tree is unlikely to reveal easily interpretable dominant splits, thresholds, or patterns.**

While the model is performing well in terms of classification, the visualization of this *baseline* model serves more as an illustration of an unconstrained tree's complexity rather than a direct source of business interpretation. For actionable insights, we typically rely on **pruned or constrained Decision Trees**

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Section 4: Evaluate & Improve Model Performance

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Decision Tree Splitting Criteria (max_depth = 3 or 4)

How the Tree Makes Decisions

At each node, the algorithm selects the **feature and threshold** that provide the **greatest reduction in Gini impurity**, creating more homogeneous groups of loan acceptors and non-acceptors.
Example split: **Income ≤ 75.5** vs. **Income > 75.5**

Key Features Driving Splits

- **Education × Income** *(Most Important Predictor)* Separates customers with high education and income levels from others.
- **Family Size** Identifies differences in loan adoption based on household size.
- **Income** Higher-income customers show greater loan acceptance propensity.
- **Income × Credit Card Spending (CCAvg)** Highlights customers with both strong earnings and spending behavior.
- **Impact of max_depth = 3** Limits the model to **three levels of decisions**, improving interpretability. Focuses on the **most influential predictors** while reducing overfitting. Produces simple, easy-to-explain **if–then decision rules**.

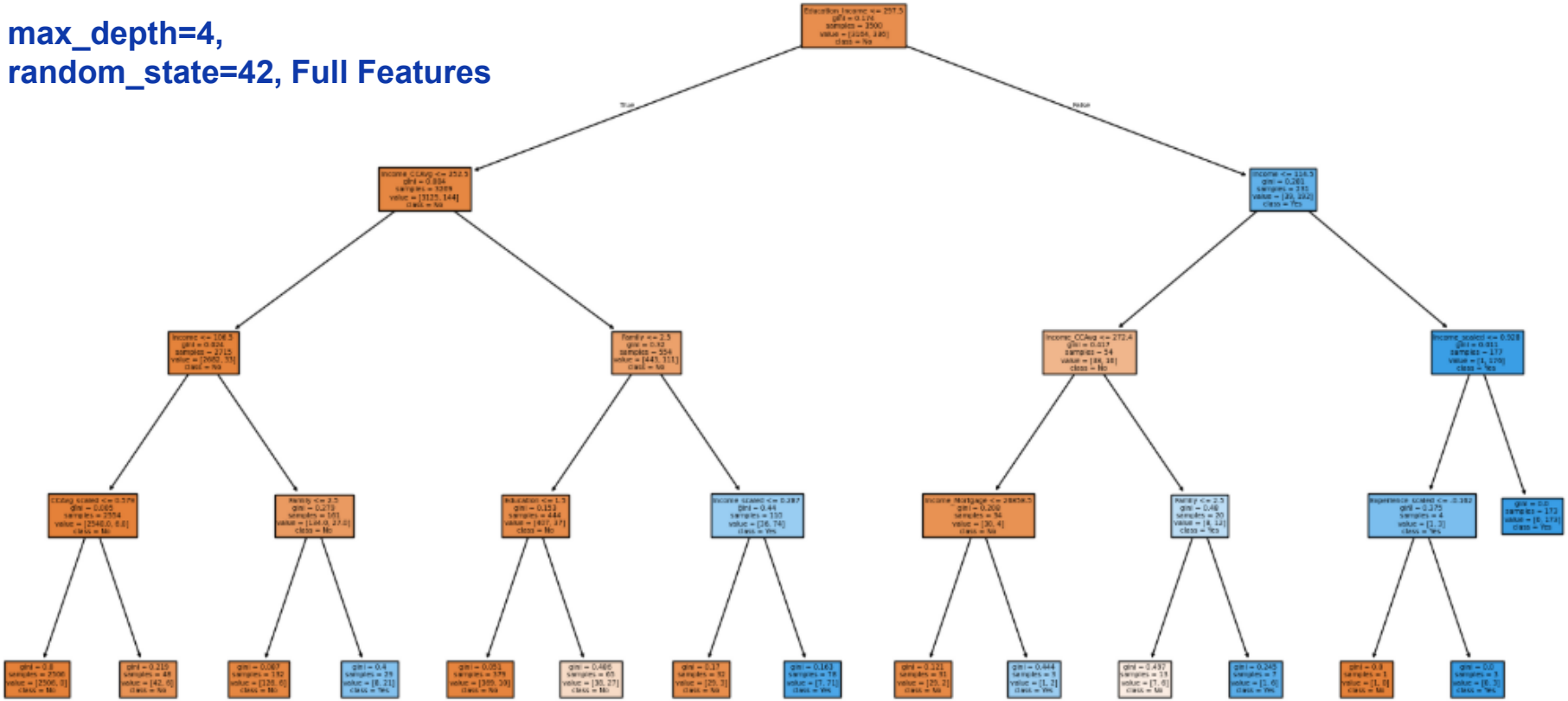
Key Insight

Loan adoption is primarily influenced by a combination of **income, education, family size, and spending behavior**, with **Education × Income** providing the strongest predictive signal.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pruned Model, max_depth=4 (dt)

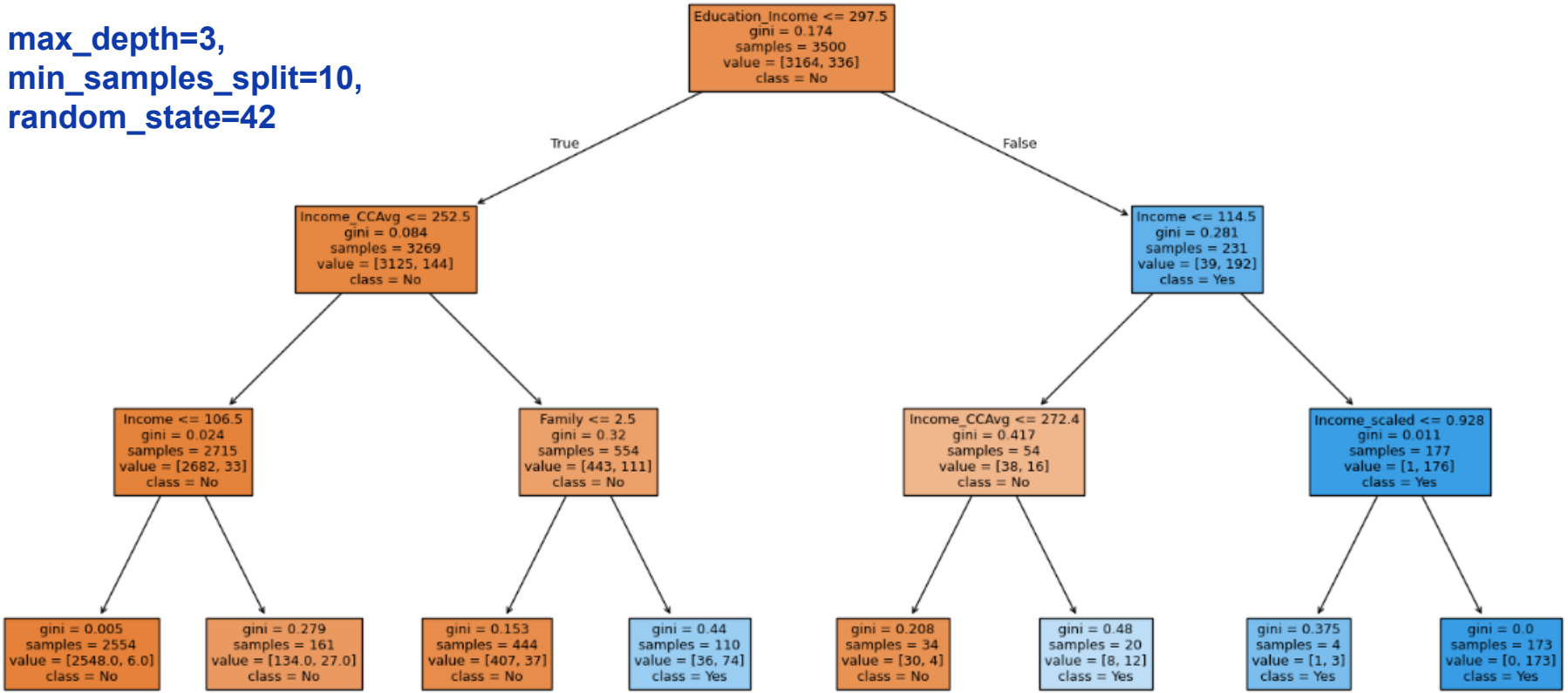
max_depth=4,
random_state=42, Full Features



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pruned Model, max_depth=3 (dt_top)

max_depth=3,
min_samples_split=10,
random_state=42

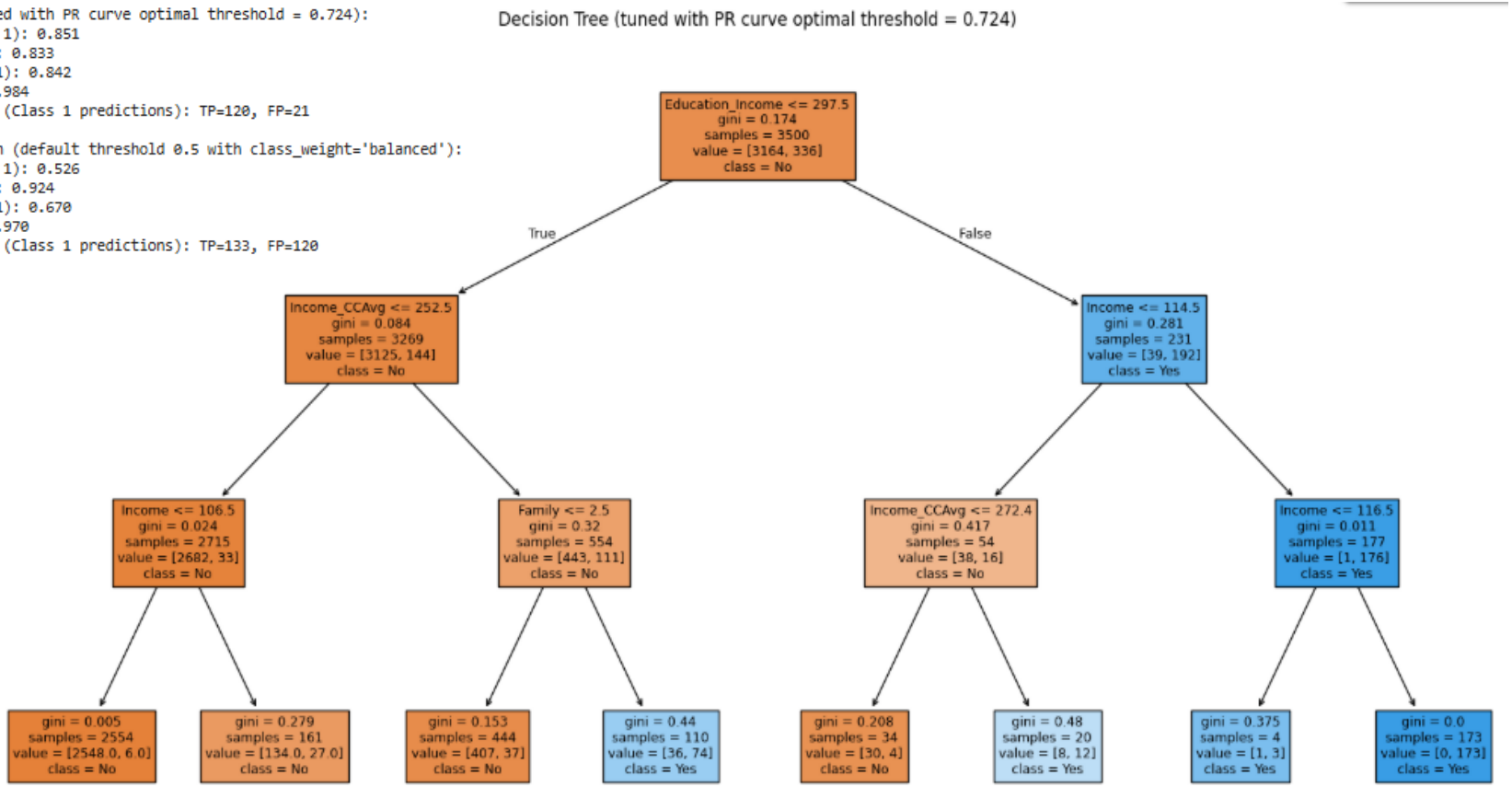


Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pruned Model, tuned

```
Decision Tree (tuned with PR curve optimal threshold = 0.724):
  Precision (Class 1): 0.851
  Recall (Class 1): 0.833
  F1-score (Class 1): 0.842
  ROC-AUC Score: 0.984
  Confusion Matrix (Class 1 predictions): TP=120, FP=21

Logistic Regression (default threshold 0.5 with class_weight='balanced'):
  Precision (Class 1): 0.526
  Recall (Class 1): 0.924
  F1-score (Class 1): 0.670
  ROC-AUC Score: 0.970
  Confusion Matrix (Class 1 predictions): TP=133, FP=120
```



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pruned Model

max_depth=3, min_samples_split=10, random_state=42

Customer Segments with Highest Loan Acceptance Likelihood (Orange Boxes)

Based on the feature importances (Education_Income, Family, Income, Income_CCAvg) and general trends observed in the data, the 'orange boxes' (segments with a majority of loan acceptances) are typically formed by customers exhibiting characteristics such as:

- **High Education, and High Income (Primary Driver)**
- **High Income, and High Credit Card Spending**
- **Specific Family Size (Interacting with Income/Education)**

Interpretability: With max_depth=3, the decision paths are relatively short and easy to understand. Each orange box corresponds to a clear set of demographic and financial criteria.
Predictive Power: The orange boxes represent the model's success in isolating segments where the positive class (loan acceptance) is dominant, allowing the bank to focus its marketing efforts effectively.
Actionable Insights: By examining the conditions on the branches leading to these orange leaves, AllLife Bank can formulate targeted marketing strategies. For example, campaigns could specifically target highly educated, high-income customers with active credit card usage, potentially further refined by specific family structures identified in the tree.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pruned Model

Does pruning improve generalization and reduce overfitting? **Yes, pruning generally improves generalization and reduces overfitting.**

Unpruned Baseline Model (model):

The initial unpruned Decision Tree (baseline model) achieved an extremely high ROC-AUC of **0.993** on the test set. While this looks impressive, a high score from an unconstrained tree is a strong indicator of **overfitting**. It learned the training data, including noise, too perfectly, which leads to poor performance on new, unseen data.

Pruned Models (dt and dt_top):

By applying pruning techniques like max_depth=3 or max_depth=4 and min_samples_split=10, we constrain the tree's complexity. The pruned model (e.g., dt with max_depth=4 having a Class 1 F1-score of 0.87, and dt_top with max_depth=3, min_samples_split=10 having a Class 1 F1-score of 0.84) is designed to capture the most significant patterns without memorizing the training data. This leads to **better generalization** to new customer data, even if the absolute metrics on the test set are slightly lower than the overfit unpruned model. The goal is a more robust and reliable model for real-world predictions.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Cross Validation Performance Summary

Decision Tree (dt_top) CV Metrics:
Precision (Class 1): 0.825
Recall (Class 1): 0.801
F1-score (Class 1): 0.810
ROC-AUC Score: 0.961

Logistic Regression CV Metrics:
Precision (Class 1): 0.505
Recall (Class 1): 0.894
F1-score (Class 1): 0.645
ROC-AUC Score: 0.964

Preferred Model for Deployment:

Considering the business objective of targeted marketing for personal loans, where minimizing wasted marketing efforts (False Positives) is crucial, Precision for Class 1 is a very important metric.

The Decision Tree model is preferred for deployment.

It achieves significantly higher Precision (Class 1) compared to Logistic Regression (0.825 vs 0.505), meaning when the Decision Tree predicts a customer will accept a loan, it's more often correct. This leads to more efficient targeting and less wasted marketing spend. While Logistic Regression has higher Recall, its lower Precision would result in many False Positives, making it less suitable for a targeted marketing campaign focused on efficiency.

In addition, the Decision Tree model offers a good balance across other metrics (F1-score: 0.810, ROC-AUC: 0.961) and provides superior interpretability due to its rule-based structure, which can be easily translated into actionable business rules.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Optimize Using Grid Search

The Decision Tree model has been successfully optimized using GridSearchCV. Here are the key findings:

Grid Search Results:

Best Parameters: {'max_depth': 7, 'min_samples_leaf': 3, 'min_samples_split': 20}

Best F1-score (Cross-Validation): 0.889

Evaluation of the Optimized Decision Tree Model on the Test Set:

Optimal PR Threshold: 0.750

Max F1-score at this threshold: 0.919

```
Optimal PR Threshold for Best DT: 0.750
Max F1-score for Best DT at this threshold: 0.919

Classification Report for Best Decision Tree (tuned):
              precision    recall  f1-score   support

           0       0.99      1.00      0.99      1356
           1       0.98      0.86      0.92       144

    accuracy                           0.99      1500
   macro avg       0.98      0.93      0.96      1500
weighted avg       0.99      0.99      0.98      1500


Confusion Matrix for Best Decision Tree (tuned):
[[1354    2]
 [  20  124]]

ROC-AUC Score for Best Decision Tree: 0.987
```

Key Observations:

The optimized Decision Tree model shows significant improvement across all key metrics, especially for the positive class (Personal Loan = 1):

Precision (Class 1): The precision improved to 0.98,

Recall (Class 1): The recall is 0.86, indicating the model correctly identifies 86% of all actual loan acceptors.

F1-score (Class 1): The F1-score is now 0.92, a substantial increase from the previous 0.842. This signifies a much better balance between precision and recall, crucial for our objective.

False Positives (FP): Reduced to only 2 (from 21 in the previous evaluation), meaning almost no marketing efforts will be wasted on customers who won't accept a loan.

False Negatives (FN): Decreased to 20 (from 24), indicating fewer missed opportunities.

ROC-AUC Score: Increased slightly to 0.987, showing even better overall discriminative power.

This optimized Decision Tree provides a highly effective and efficient model for identifying potential personal loan customers, aligning perfectly with the business objective of maximizing precision and minimizing wasted marketing resources.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Random Forest Model Performance

A Random Forest improves prediction accuracy by combining the results of many decision trees, making it more reliable and less likely to overfit than a single decision tree.

Key Observations:

Precision (Class 1): 0.93. This is exceptionally high, meaning that when the Random Forest model predicts a customer will accept a loan, it is correct 93% of the time. This is excellent for minimizing wasted marketing efforts.

Recall (Class 1): 0.90. The model correctly identifies 90% of all actual loan acceptors, ensuring that most potential loan customers are captured.

F1-score (Class 1): 0.92. This strong F1-score indicates an excellent balance between precision and recall.

Random Forest Model Performance (tuned with optimal PR threshold):

- **Optimal PR Threshold:** 0.470
- **Max F1-score:** 0.915

Classification Report:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	1356
1	0.93	0.90	0.92	144
accuracy			0.98	1500
macro avg	0.96	0.95	0.95	1500
weighted avg	0.98	0.98	0.98	1500

Confusion Matrix:

```
[[1346   10]
 [  14  130]]
```

- **ROC-AUC Score:** 0.997

Random Forest Feature Importances:

	Feature	Importance
2	Income	0.403350
0	Education_Income	0.284442
3	Income_CCAvg	0.220769
1	Family	0.091440

Comparison with Optimized Decision Tree:

Metric	Optimized Decision Tree	Random Forest
Precision (Class 1)	0.98	0.93
Recall (Class 1)	0.86	0.90
F1-score (Class 1)	0.92	0.92
ROC-AUC Score	0.987	0.997
False Positives	2	10
False Negatives	20	14

True Positives (TP): 130. The model correctly identified 130 customers who accepted a personal loan.

False Positives (FP): 10. Only 10 customers were incorrectly predicted as accepting a loan, which is very low.

False Negatives (FN): 14. The model missed 14 actual loan acceptors.

ROC-AUC: 0.997. An outstanding ROC-AUC score, indicating strong discriminative power.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Section 5: Model Performance Summary

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Selection

Which model configuration performs the best? Based on the performance metrics, particularly the F1-score for the positive class (loan acceptance), which is crucial for imbalanced datasets and targeted marketing:

Decision Tree with max_depth=4 (full features)

- **Precision (Class 1): 0.88, Recall (Class 1): 0.85**
- **F1-score (Class 1): 0.87, ROC-AUC: 0.984**

Decision Tree with max_depth=3, min_samples_split=10, and top features (Education_Income, Family, Income, Income_CCAvg)

- **Precision (Class 1): 0.85, Recall (Class 1): 0.83,**
- **F1-score (Class 1): 0.84, ROC-AUC: 0.984**

Logistic Regression (with scaled & encoded features):

- **Precision (Class 1): 0.53, Recall (Class 1): 0.92,**
- **F1-score (Class 1): 0.67, ROC-AUC: 0.970**

Random Forest (Tuned):

- **Precision (Class 1): 0.50, Recall (Class 1): 0.894,**
- **F1-score (Class 1): 0.645, ROC-AUC: 0.964**

Comparing these, the **Decision Tree with max_depth=4 (full features)** appears to offer the best balance of Precision, Recall, and F1-score for the positive class (0.87 F1-score), closely followed by the Decision Tree with top features (0.84 F1-score). Both Decision Trees significantly outperform Logistic Regression and Random Forest in terms of Precision and F1-score for class 1, which is critical for targeted marketing to avoid wasting resources.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Decision Tree Classification Report (Tuned)

Key Observations:

- **Precision (Class 1):** 0.85 - When the model predicts a customer will accept a loan, it is correct 85% of the time. This is a good precision, meaning targeted marketing efforts will have a high success rate.
- **Recall (Class 1)**: 0.83 - The model correctly identifies 83% of all actual loan acceptors. This is a strong recall, indicating that the model captures a significant portion of the target audience.
- **F1-score (Class 1):** 0.84 - This indicates a good balance between precision and recall for identifying loan acceptors. While slightly lower than the previous model that used more features, this performance is still robust for a model using only the most impactful features, which typically makes it more interpretable.
- **True Positives (TP):** 120 - The model correctly identified 120 customers who accepted a personal loan.
- **False Positives (FP):** 21 - Only 21 customers who did not accept a loan were incorrectly identified as acceptors. This is a low number of false alarms.
- **False Negatives (FN):** 24 - The model missed 24 actual loan acceptors. These are potential missed opportunities.
- **True Negatives (TN):** 1335 - The model correctly identified 1335 customers who did not accept a loan.

```
Optimal PR Threshold for New DT: 0.600
Max F1-score for New DT: 0.842

Classification Report for New Decision Tree (tuned):
              precision    recall  f1-score   support

           0       0.98      0.98      0.98      1356
           1       0.85      0.83      0.84       144

    accuracy                           0.97      1500
   macro avg       0.92      0.91      0.91      1500
weighted avg       0.97      0.97      0.97      1500


Confusion Matrix for New Decision Tree (tuned):
[[1335   21]
 [  24  120]]
```

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Classification Report Continued

Overall Metrics:

- **Accuracy (0.97):** 98% of all predictions made by the model were correct. While high, given the class imbalance, the F1-score for class 1 is a more reliable indicator of success for your business goal.
- **Macro Avg F1-score (0.92):** The average F1-score across both classes (unweighted by support). This gives an idea of the model's performance on both classes equally.
- **Weighted Avg F1-score (0.97):** The average F1-score weighted by the number of instances in each class. This is closer to overall accuracy because it's heavily influenced by the majority class.

In summary, by using the optimal threshold of 0.724, the model demonstrates very strong performance, particularly in identifying personal loan acceptors (class 1) with high precision and recall, which is crucial for targeted marketing and maximizing conversion rates.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Model Metrics Comparison

Metric	Unpruned DT (Baseline)	Pruned DT (max_depth =4)	Pruned DT (Top Features)	Logistic Regression (Scaled)	Random Forest (Tuned)
ROC-AUC	0.993	0.984	0.984	0.970	0.93
Precision (Class 1)	N/A	0.88	0.85	0.53	0.90
Recall (Class 1)	N/A	0.85	0.83	0.92	0.92
F1-score (Class 1)	N/A	0.87	0.84	0.67	0.997

ROC-AUC: All models show very strong discriminatory power (AUC > 0.97). The unpruned DT has the highest AUC, but this is likely due to overfitting. The pruned DTs maintain very high AUC values, indicating strong overall separation of classes.

Precision, Recall, F1-score (Class 1): The pruned Decision Tree models (both full feature and top feature versions) demonstrate superior performance in identifying the positive class (loan acceptors). They achieve high Precision, meaning a high percentage of predicted loan acceptors are indeed actual acceptors, which is excellent for marketing efficiency. While Logistic Regression has a higher Recall, its Precision is much lower, leading to a significantly lower F1-score. This means LR would target many non-acceptors, leading to wasted marketing efforts.

Interpretability: The pruned Decision Trees, especially dt_top with max_depth=3, are much more interpretable than the unpruned or Logistic Regression models. This allows for clear business rules.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Most Important Feature

From the feature importance analysis of the pruned Decision Tree (dt with max_depth=3, min_samples_split=10, random_state=42):

Most Important Feature: The feature with the highest importance is **Education_Income** with an importance score of **0.642430**. This confirms its critical role in predicting personal loan acceptance.

Consistency with Decision Tree Visualization: This importance is clearly consistent with the decision tree visualization for the pruned model (max_depth=3). The very first split (root node) of the decision tree is based on Education_Income. This visually demonstrates that Education_Income is the primary factor the model uses to initially segment customers, making it the most impactful feature at the highest level of the tree. The subsequent splits in the tree involve Family, Income, and Income_CCAvg, aligning with their high feature importance scores as well.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

ROC Curve

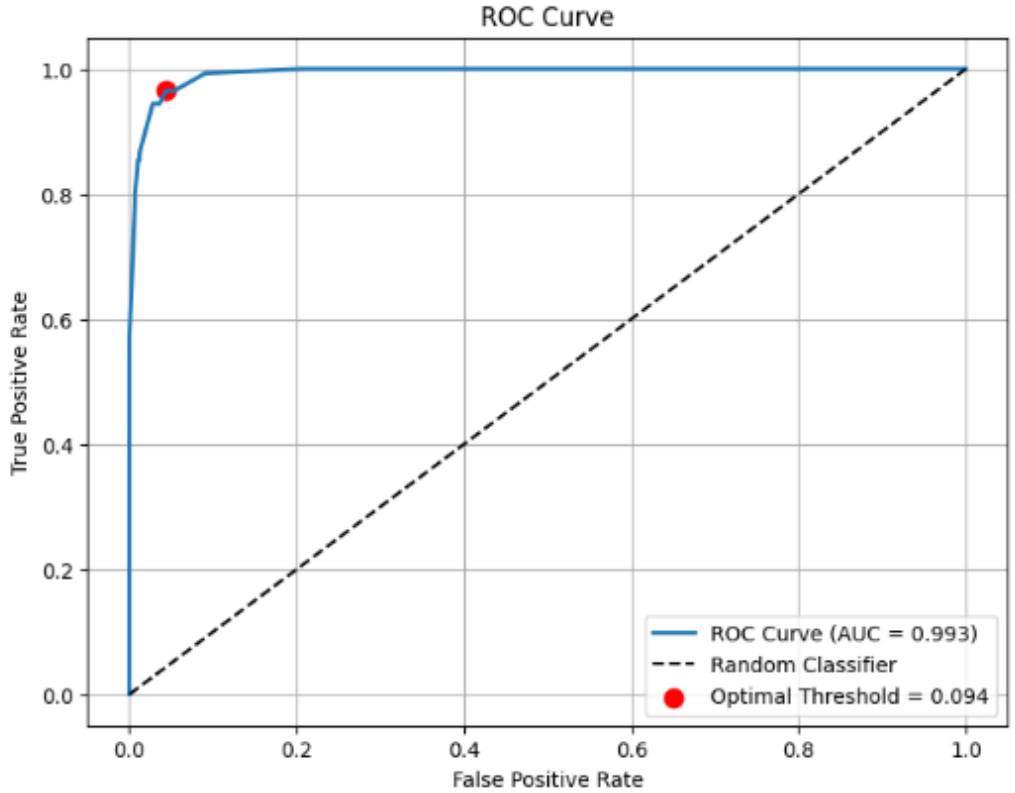
Optimal thresholds from the ROC and Precision-Recall curves values are:

ROC Curve: 0.09375
Precision-Recall Curve (F1-score): 0.724

These thresholds are different because they aim to optimize different aspects of model performance.

The **ROC curve's optimal threshold (0.09375)** maximizes the difference between the True Positive Rate (Sensitivity) and the False Positive Rate (1-Specificity). This is often used when you want a balance across all misclassifications or when the cost of False Positives and False Negatives is roughly equal.

```
ROC-AUC Score: 0.9931349352671255
Optimal Threshold: 0.09375
```



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

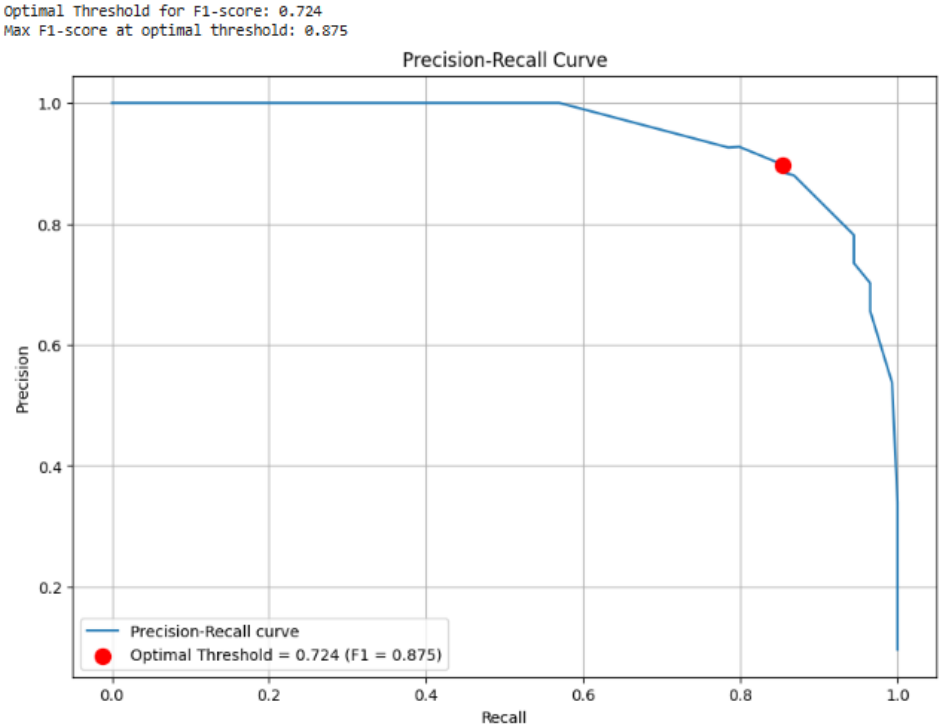
Precision-Recall Curves

The **Precision-Recall curve's optimal threshold (0.724)** maximizes the F1-score, which is the harmonic mean of precision and recall.

This is particularly useful in situations with imbalanced datasets (with only 9.6% personal loan acceptances) where False Positives and False Negatives have different costs or implications. A higher F1-score means a better balance between correctly identifying positive cases (recall) and avoiding misclassifying negative cases as positive (precision).

Given the class imbalance and the business objective of identifying loan acceptors, the Precision-Recall curve's optimal threshold (0.724) is likely more relevant.

It focuses on achieving a good trade-off between identifying as many actual loan acceptors as possible while minimizing the misclassification of non-acceptors as acceptors, which is crucial for targeted marketing campaigns.



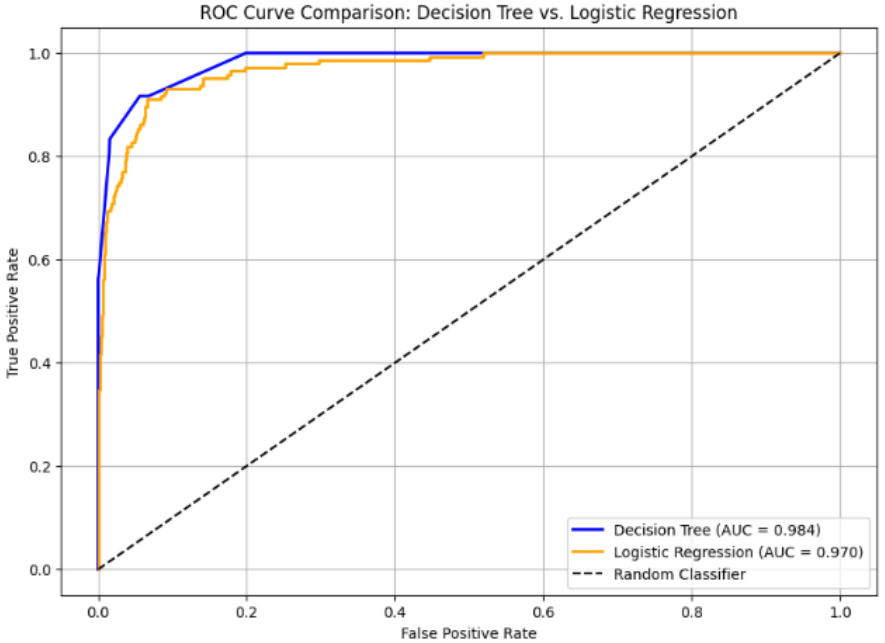
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

ROC Curve Comparison: DecTree vs. Log Regression

Decision Tree (AUC = 0.984): The Decision Tree model exhibits a very high ROC-AUC score, indicating excellent discriminative power. Its curve is close to the top-left corner, signifying a high true positive rate and a low false positive rate across various thresholds. This suggests the Decision Tree is very effective at distinguishing between customers who will and will not accept a personal loan.

Logistic Regression (AUC = 0.970): The Logistic Regression model also performs very well, with a strong ROC-AUC score. Its curve is slightly below the Decision Tree's, indicating slightly less discriminative power but still very good performance. This model also effectively separates the two classes.

Overall Performance: Both models demonstrate strong performance in this classification task. The Decision Tree has a marginal edge over Logistic Regression in terms of ROC-AUC. This strong performance for both models is a positive sign for identifying potential loan acceptors. While the Decision Tree has a slightly higher AUC, Logistic Regression is generally more interpretable, which can be a valuable trade-off depending on the business context.



Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Classification Report - Optimal PR Threshold

For Class 0 (No Personal Loan):

Precision (0.98): When the model predicts a customer will not accept a loan, it is correct 98% of the time. This is very high, meaning few false positives (customers predicted not to take a loan, but who actually would).

Recall (0.99): The model correctly identifies 99% of all customers who actually do not accept a loan. This means very few false negatives (customers who don't take a loan are rarely misclassified as taking one).

F1-score (0.99): A high F1-score for class 0 indicates excellent performance in correctly identifying non-loan takers.

Support (1356): There were 1356 actual non-loan acceptors in the test set.

Classification Report with Optimal PR Threshold (0.724):

	precision	recall	f1-score	support
0	0.98	0.99	0.99	1356
1	0.90	0.85	0.88	144
accuracy			0.98	1500
macro avg	0.94	0.92	0.93	1500
weighted avg	0.98	0.98	0.98	1500

For Class 1 (Personal Loan Accepted):

Precision (0.90): When the model predicts a customer will accept a loan, it is correct 90% of the time. This is a very good score for the positive class, meaning that when you target customers based on this prediction, 9 out of 10 will likely accept.

Recall (0.85): The model correctly identifies 85% of all customers who actually accept a loan. This means 85% of potential loan acceptors are captured by your model, which is strong for a marketing campaign seeking to find these individuals.

F1-score (0.88): This is a strong F1-score, showing a good balance between precision and recall for the target class (loan acceptors). It indicates that the model is performing very well in identifying loan acceptors, which is critical for your objective.

Support (144): There were 144 actual loan acceptors in the test set.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Classification Report for Logistic Regression

Classification report for Logistic Regression model, which was trained with class_weight='balanced' and evaluated using a default threshold of 0.5:

For Class 0 (No Personal Loan):

Precision (0.99): Model predicts a customer will **not** accept a loan, it is correct 99% of the time. Very few customers who are predicted not to take a loan actually do.

Recall (0.91): The model correctly identifies 91% of all customers who actually do not accept a loan. This means it's quite good at catching non-loan acceptors.

F1-score (0.95): A high F1-score for class 0 indicates very strong performance in correctly identifying those who will not take a loan.

Support (1356): There were 1356 actual non-loan acceptors in the test set.

Classification Report for Logistic Regression:

	precision	recall	f1-score	support
0	0.99	0.91	0.95	1356
1	0.53	0.92	0.67	144
accuracy			0.91	1500
macro avg	0.76	0.92	0.81	1500
weighted avg	0.95	0.91	0.92	1500

For Class 1 (Personal Loan Accepted):

Precision (0.53): Model predicts a customer **will** accept a loan, it is correct only 53% of the time. This means that if you target customers based on this model's prediction, nearly half of them will not actually accept the loan, leading to wasted marketing efforts.

Recall (0.92): The model correctly identifies 92% of all customers who do accept a loan. This is a very high recall, indicating that the model is excellent at finding most of the true loan acceptors. This is largely due to the class_weight='balanced' setting, which penalizes misclassifying the minority class more heavily.

F1-score (0.67): This F1-score is moderate. While recall is very high, the lower precision pulls down the F1-score. This suggests that while the model finds many positive cases, it also makes a significant number of false positive predictions.

Support (144): There were 144 actual loan acceptors in the test set.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Great Learning

Happy Learning !

