Personal Loan Campaign - Problem Statement

**Submission type**

:

File Upload

**Due Date**

:

Dec 07, 4:30 AM CST

**Total Marks**

:

60

**Available from**

:

Nov 07, 8:30 AM

**Description**

**Background and Context**

AllLife Bank is a US bank that has a growing customer base. The majority of these customers are liability customers (depositors) with varying sizes of deposits. The number of customers who are also borrowers (asset customers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process, earn more through the interest on loans. In particular, the management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors).

A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio.

You as a Data scientist at AllLife bank have to build a model that will help the marketing department to identify the potential customers who have a higher probability of purchasing the loan.

**Objective**

To predict whether a liability customer will buy personal loans, to understand which customer attributes are most significant in driving purchases, and to identify which segment of customers to target more.

**Data Dictionary**  
\* ID: Customer ID  
\* Age: Customer’s age in completed years  
\* Experience: #years of professional experience  
\* Income: Annual income of the customer (in thousand dollars)  
\* ZIP Code: Home Address ZIP code.  
\* Family: the Family size of the customer  
\* CCAvg: Average spending on credit cards per month (in thousand dollars)  
\* Education: Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional  
\* Mortgage: Value of house mortgage if any. (in thousand dollars)  
\* Personal\_Loan: Did this customer accept the personal loan offered in the last campaign?  
\* Securities\_Account: Does the customer have securities account with the bank?  
\* CD\_Account: Does the customer have a certificate of deposit (CD) account with the bank?  
\* Online: Do customers use internet banking facilities?  
\* CreditCard: Does the customer use a credit card issued by any other Bank (excluding All life Bank)?

**Submission Guidelines**

1. There are two ways to work on this project:

**i. Full-code way:**The full code way is to write the solution code from scratch and only submit a final Jupyter notebook with all the insights and observations.

**ii. Low-code way:**The low-code way is to use an existing solution notebook template to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their individual learning aspirations and outcomes. The below table elaborates on these two options.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Submission type | Who should choose | What is the same across the two | What is different across the two | Final submission file [IMP] | Submission Format |
| Full-code | Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch | Perform exploratory data analysis to identify insights and recommendations for the problem | Focus on code writing: 10-20% grading on the quality of the final code submitted | Solution notebook from the full-code template submitted in .html format | .html |
| Low-code | Learners who aspire to be in managerial roles in the future-focussed on solution review, interpretation, recommendations, and communicating with business |  | Focus on business presentation: 10-20% grading on the quality of the final business presentation submitted | Business presentation in .pdf format with problem definition, insights, and recommendations | .pdf |

Please follow the below steps to complete the assessment. **Kindly note that if you submit a presentation along with the notebook, ONLY the presentation will be evaluated.** Please make sure that all the sections mentioned in the rubric have been covered in your submission.

**i. Full-code version**

* Download the full-code version of the learner notebook.
* Follow the instructions provided in the notebook to complete the project.
* Clearly write down insights and recommendations for the business problems in the comments.
* Submit only the solution notebook prepared from the learner notebook [format: .html]

**ii. Low-code version**

* Download the low-code version of the learner notebook.
* Follow the instructions provided in the notebook to complete the project.
* Prepare a business presentation with insights and recommendations to the business problem.
* Submit only the presentation [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated if

* it is submitted post-deadline, or,
* more than 1 file is submitted.

**Best Practices for Full-code submissions**

* The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
* The notebook should be run from start to finish in a sequential manner before submission.
* It is important to remove all warnings and errors before submission.
* The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
* Please refer to the FAQ page for common project-related queries.

**Best Practices for Low-code submissions**

* The presentation should be made keeping in mind that the audience will be the Data Science lead of a company.
* The key points in the presentation should be the following:
  + Business Overview of the problem and solution approach
  + Key findings and insights which can drive business decisions
  + Business recommendations
  + Focus on explaining the key takeaways in an easy-to-understand manner.
  + The inclusion of the potential benefits of implementing the solution will give you the edge.
* Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
* The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
* Please refer to the FAQ page for common project-related queries.

Happy Learning!

**Scoring guide (Rubric) - Personal Loan Campaign Rubric**

| **Criteria** | **Points** |
| --- | --- |
| **Define the problem and perform an Exploratory Data Analysis**  - Problem definition, questions to be answered - Data background and contents - Univariate analysis - Bivariate analysis - Key meaningful observations for each of the plot | 10 |
| **Data pre-processing**  Prepare the data for analysis: - Missing Value Treatment (if needed) - Outlier Detection (treat, if needed) - Feature Engineering - Data preparation for modelling | 6 |
| **Model building - Decision Tree**  - Define model evaluation criterion - Build the model and comment on the model performance. - Visualize the decision rules and important features | 10 |
| **Model Performance Evaluation and Improvement**  - Try and improve the model performance by pruning (Both Post and Pre pruning) - Check the performance of the pruned models, compare the performance of all the models built, and select the final model - Find the decision rules and check feature importance for the final model | 20 |
| **Actionable Insights & Recommendations**  - Conclude with the key takeaways for the marketing team - What would your advice be on how to do this campaign? | 6 |
| **Presentation/Notebook - Overall Quality**  - Structure and flow - Crispness - Visual appeal - Conclusion and Business Recommendations OR - Structure and flow - Well-commented code - Conclusion and Business Recommendations | 8 |
| Points | 60 |

FAQ - Personal Loan Campaign

#### ****1. How should one approach the Personal Loan Campaign project?****

* Before starting the project, please read the problem statement carefully and go through the criteria and descriptions mentioned in the rubric.
* Once you understand the task, download the dataset and import it into a Jupyter notebook to get started with the project.
* To work on the project, you should start with data preprocessing and EDA using descriptive statistics and visualizations.
* Once the EDA is completed and data is preprocessed, you can use the data to build a model and check its performance.
* It is important to close the analysis with key findings and recommendations to the business.

#### ****2. Decision Tree arrows are missing, how to fix this?****

Use the following code as a reference to resolve the issue and make necessary changes (name of the model, feature names, etc):

plt.figure(figsize=(20,30))

out = tree.plot\_tree(model,feature\_names=feature\_names,filled=True,fontsize=9,node\_ids=False,class\_names=None,)

#below code will add arrows to the decision tree split if they are missing

for o in out:

arrow = o.arrow\_patch

if arrow is not None:

arrow.set\_edgecolor('black')

arrow.set\_linewidth(1)

plt.show()

#### ****3.**** ****How to deal with "ZIPCode'' as it is a numeric value but it's also essentially a category?****

You can explore the following links to deal with zip codes:

1. [uszipcode](https://pypi.org/project/uszipcode/" \t "_blank) - Python package that can help in mapping zip codes to different locations

2. <https://www.smartystreets.com/articles/zip-4-code> - Description of how zip codes are created in the US.

#### ****4. Should I create dummies for the columns that only have 0’s and 1’s?****

#### No, it is not necessary to create dummies for these columns. ****5. I'm trying to post-prune the decision tree.  But I'm getting the following error:****

“**ValueError**: ccp\_alpha must be greater than or equal to 0"

#### ****How to resolve this?****

To resolve this error kindly use absolute values (positive value) of alpha. Use the following lines of code to resolve the error:

ccp\_alphas, impurities = abs(path.ccp\_alphas), path.impurities

#### ****6. I get the following error:****

ModuleNotFoundError: No module named 'nb\_black'

#### ****how do I install nb\_black?****

Run the below code in the anaconda prompt

pip install nb-black

or run the below code in jupyter notebook.

!pip install nb-black