PROJECT BRIEF

PREDICTIVE MODELING FOR A CUSTOMIZED MARKETING CAMPAIGN: PERSONAL LOANS

Nicole Nixen

Original Completion Date: 4.16.24

INTRODUCTION

- This presentation **summarizes my capstone project** for Intro to Machine Learning for the Post-Grad Program of Business Applications of AI and Machine Learning via UT Austin/Great Learning.

- Using **Python** to create a model, this project demonstrates how predictive modeling can support targeted marketing campaign.

- For full code, notebook, and synthetic dataset: GitHub Portfolio

BUSINESS PROBLEM OVERVIEW

- A bank is exploring ways to **convert** liability customers (depositors) to asset customers (personal loan borrowers).
- The bank ran a loan campaign for liability customers. **9% of the targeted customers converted** to an asset customer by accepting a personal loan.
- The bank wants to **improve the conversion rate** of targeted customers on the next loan campaign.
- A **Python** model was built to understand which liability customers would have a higher probability of converting to an asset customer
- The predictive model presented here **identifies depositors most likely to accept** a personal loan offer to improve marketing campaign targeting and forecasting.

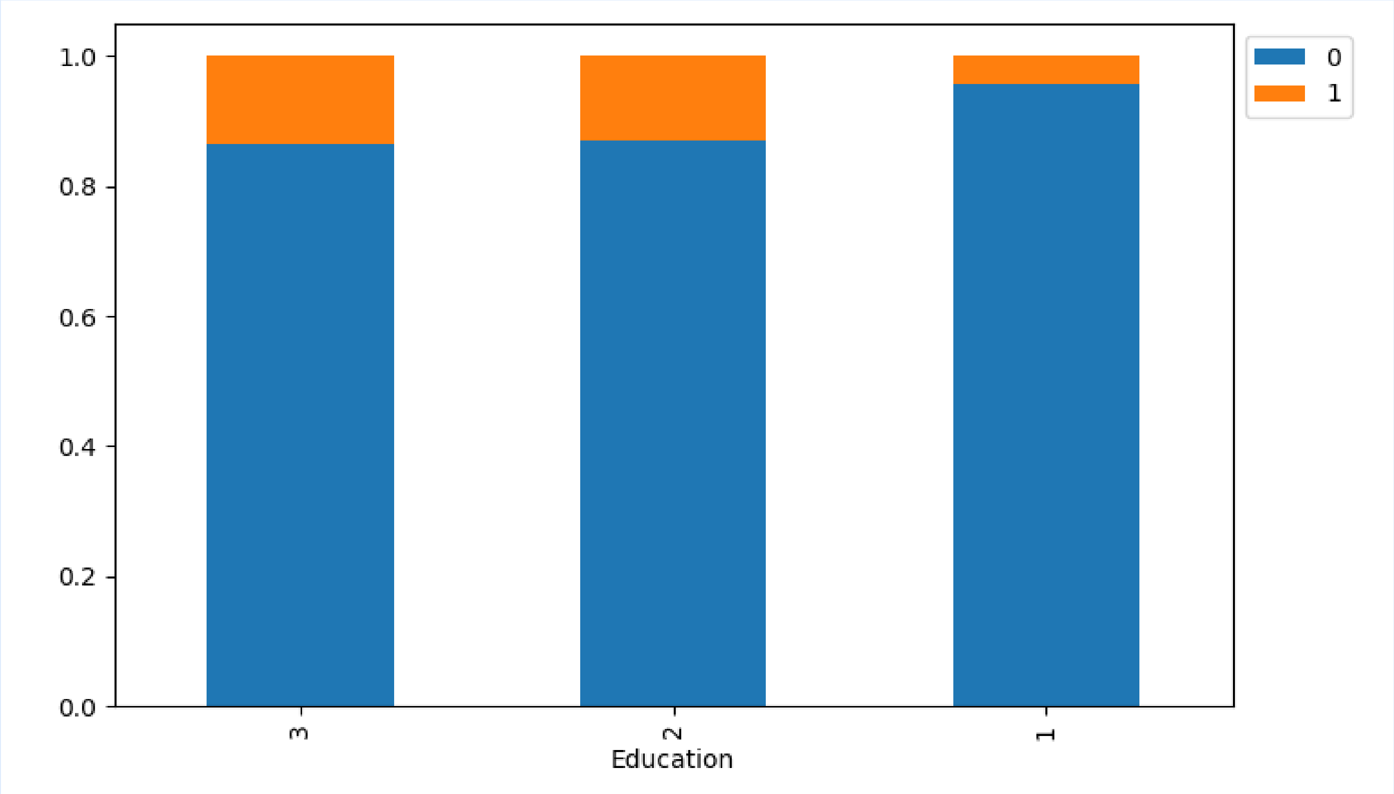
METHODOLOGY

- The dataset was provide by edtech partner Great Learning, and originally obtained through a publicly available dataset on Kaggle.
- The dataset used in this project does not contain personally identifiable information and represents the previous marketing campaign that netted a 9% conversion rate.
- An exploratory data analysis (EDA) was conducted to find trends and correlations in the data.
  - Univariate analysis examined insights on skewness, IQR and outliers of age, education, income, credit card average, mortgage, family, security account, CD account, online, credit card and zip code.
  - Bivariate analysis looked for correlations among the combinations of age, education, income, credit card average, mortgage, family, security account, CD account, online, credit card and zip code.

METHODOLOGY

- One result of the EDA revealed that the field Experience was perfectly correlated with the field Age; as a result the field Experience was dropped from the dataset.
- Next, a decision tree model was built using Python.
- The decision tree classifier used a 70/30 train/test split, and the selected model prioritized **Recall.**
- Further modifications were run using pre and post pruning methods, but ultimately the initial model delivered the most accurate results in predicting whether or not a customer would accept a loan.

KEY INSIGHTS FROM EDA



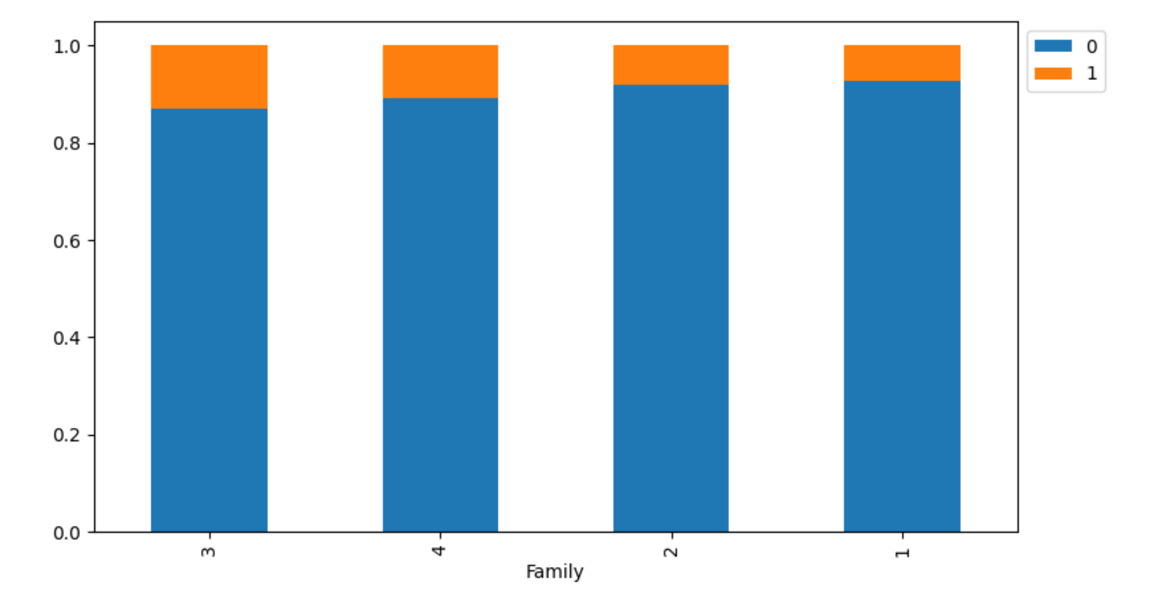
Customers with higher levels of education were more likely to accept a loan.

Education Level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional

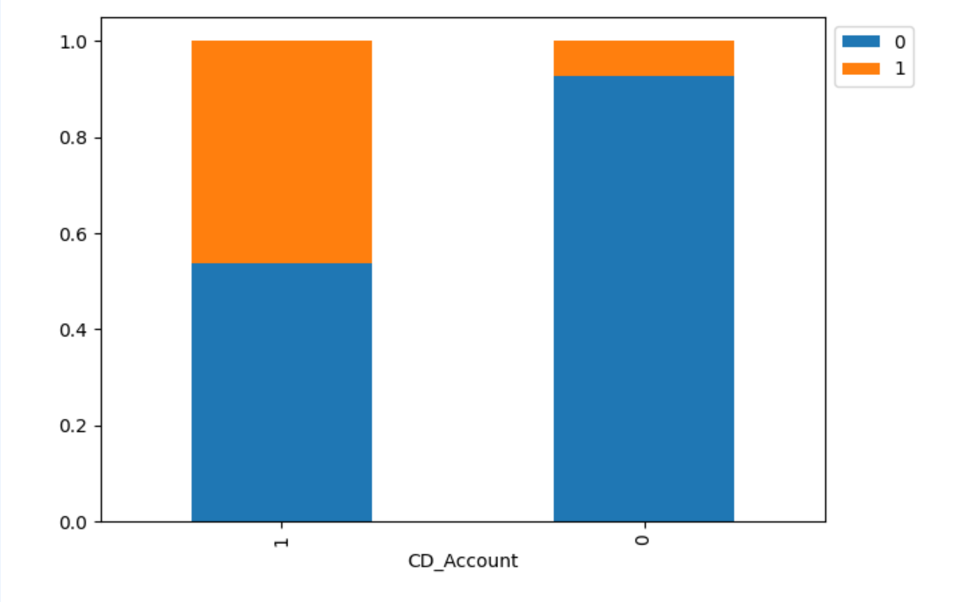
Blue (0): Did not accept a loan   Orange (1): Accepted a loan

KEY INSIGHTS FROM EDA

Family Size (1-4 people)



CD_Account (does customer have a CD with the bank) 0: No 1: Yes



Blue (0): Did not accept a loan   Orange (1): Accepted a loan

Customers with a family size of 3-4 were more likely to accept a loan.
Customers with a CD account with the bank were more likely to accept a loan.

MODEL OVERVIEW

- The data was split 70/30:
  - 70% of the data was assigned to the training data
  - 30% was assigned to the test data.
- Both the training and test data contained 91% of customers who declined the loan, and 9% who accepted the loan.
- This mirrors the overall loan acceptance rate of 9% in total observations.
- Both false positives and false negatives would result in lost revenue opportunity for the bank, but false negatives would result in greater lost opportunity (i.e. predicting a customer would not open a loan when in reality they would).
- To minimize false negatives, **Recall** was the main criteria in choosing the best mode.
- The recall score of the recommended model was 93%.

	Accuracy	Recall	Precision	F1
0	0.986	0.932886	0.926667	0.929766

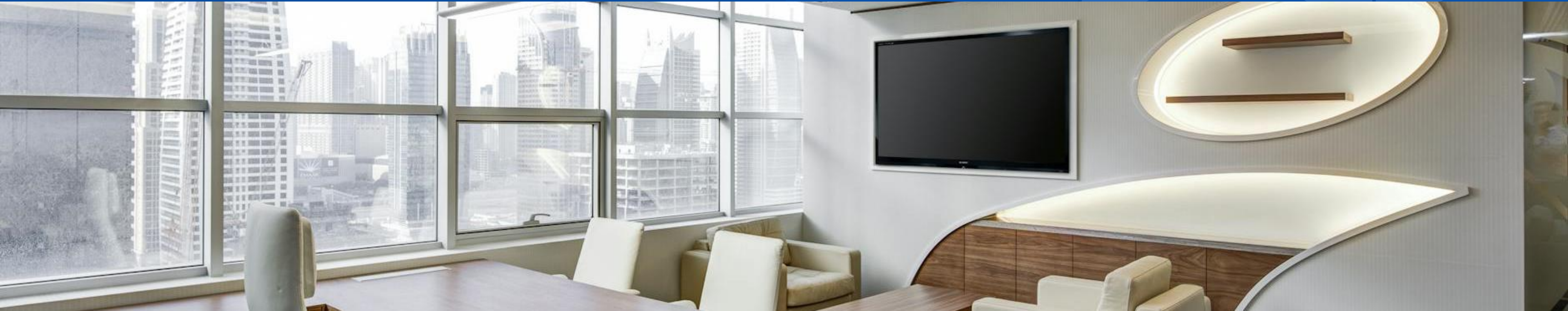
ACTIONABLE INSIGHTS AND RECOMMENDATIONS

- The bank should focus marketing efforts on loan acceptance towards customers with:

    - Higher education
    - Family size of 3-4
    - Customers who hold a CD with the bank

- A Decision Tree predictive model with minimal pruning can accurately predict which customers are likely to accept a loan.

- Future recommended analysis would be to evaluate the likelihood of loan repayment or likelihood of a customer defaulting on a loan as a modification to any algorithm predicting which customers to offer a loan.



CONTACT INFO:

NICOLE NIXEN

HTTPS://WWW.LINKEDIN.COM/IN/NICOLENIXEN/

FOR FULL CODE, NOTEBOOK, AND SYNTHETIC DATASET: GITHUB PORTFOLIO