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 and campaign forecasting in financial services.

- 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.

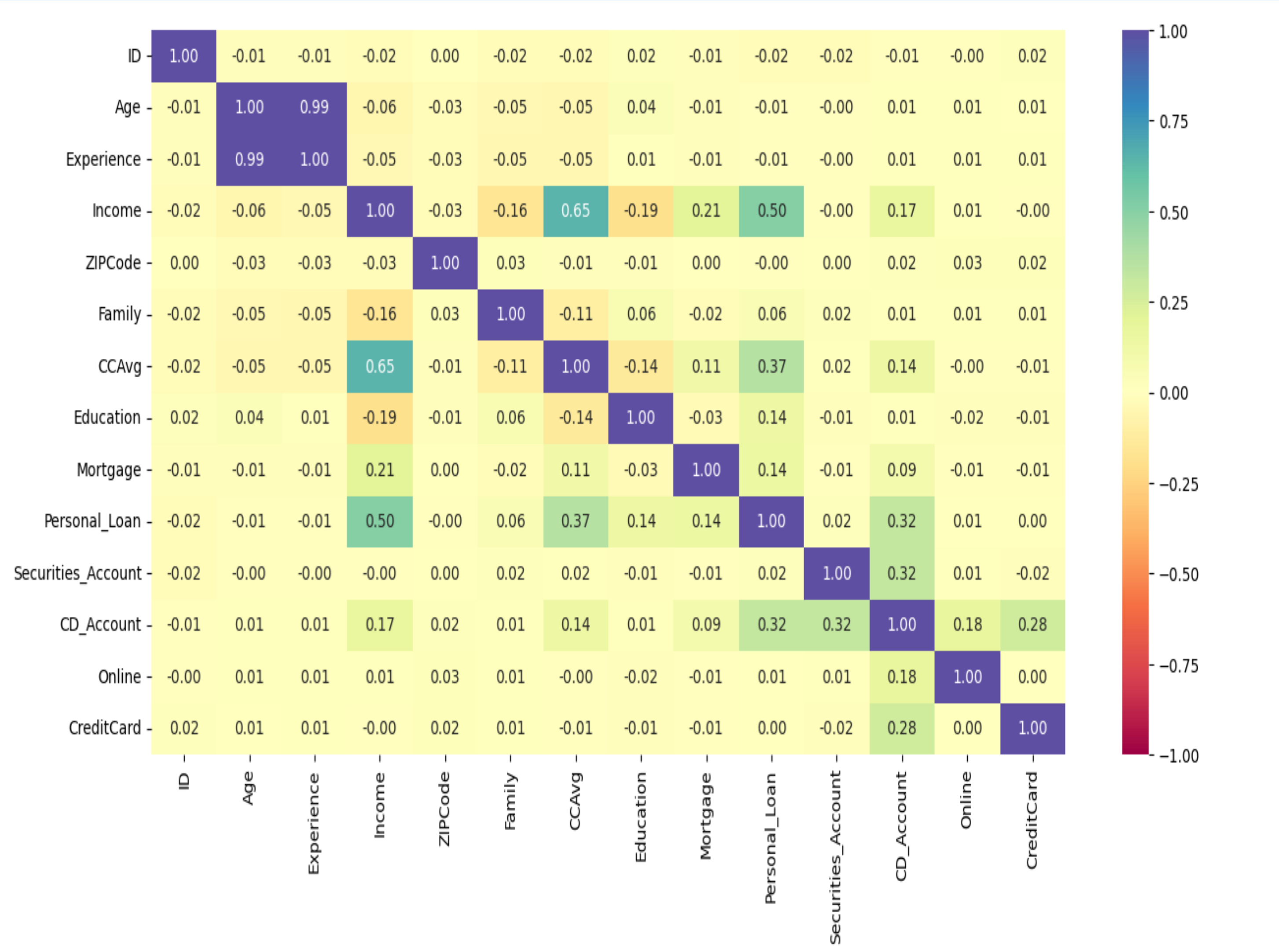
METHODOLOGY

- 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



HEATMAP GENERATED FROM DATA

- Age and experience are perfectly correlated.
- Customers with higher levels of education were more likely to accept 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 EVALUATION CRITERIA

Potential errors that could occur with this model is:

- **False Negative:** Predicting a customer will not accept a loan, but in reality they would accept a loan offer.
- **Fales Positive:** Predicting a customer will accept a loan offer, but in reality they would not accept a loan offer.
- In this scenario, both false positives and false negatives result in lost revenue opportunity for the bank.
- The cost of predicting a customer will not accept a loan when in fact they would results in a greater loss of revenue to the bank.

For this case, **Recall** would be the best metric to use to determine the best model.

Recall looks for the true positive rate.

FINAL 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.
- Initially, the training data returned accuracy, recall, precision and F1 scores of 100%, which would indicate a perfectly predictive model.

Training performance comparison:

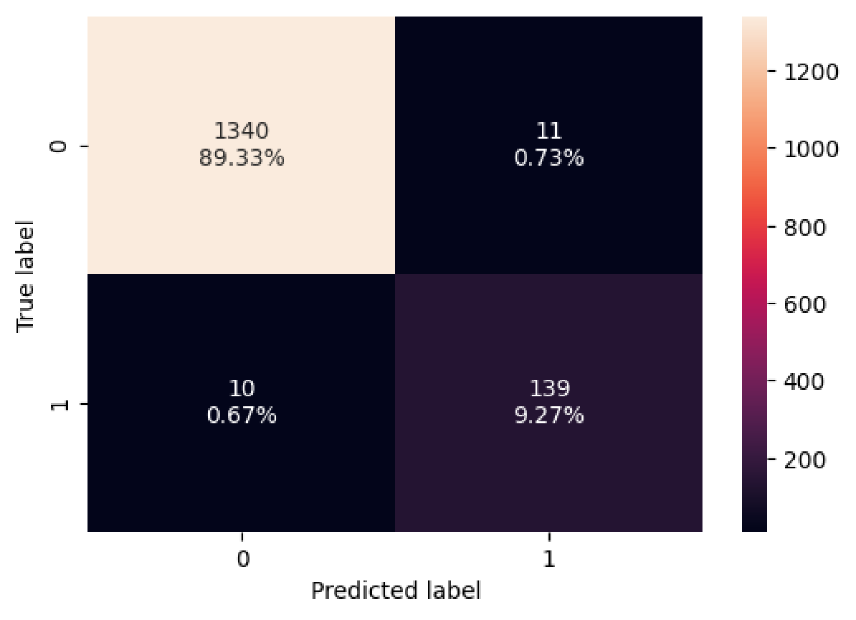
	Decision Tree sklearn	Decision Tree (Pre-Pruning)
Accuracy	1.0	0.987714
Recall	1.0	0.873112
Precision	1.0	0.996552
F1	1.0	0.930757

FINAL MODEL OVERVIEW

The test data came back with lower scores across all metrics.

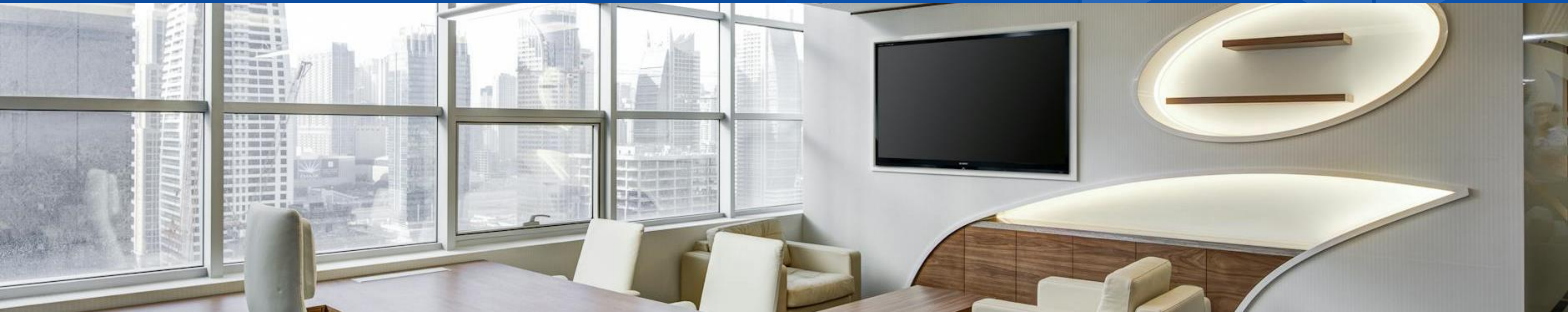
	Accuracy	Recall	Precision	F1
0	0.986	0.932886	0.926667	0.929766

- The test data results suggest the model overfit the training data.
- However, the scores from the confusion matrix for the training data indicate this is a useful model.



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
    - Customers with income below $100K
- A Decision Tree predictive model with minimal pruning can accurately predict which customer 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