



Thera Bank Customer Conversion

Converting current depositors to
depositors/borrowers

Team3 Loaners

Gary Gao

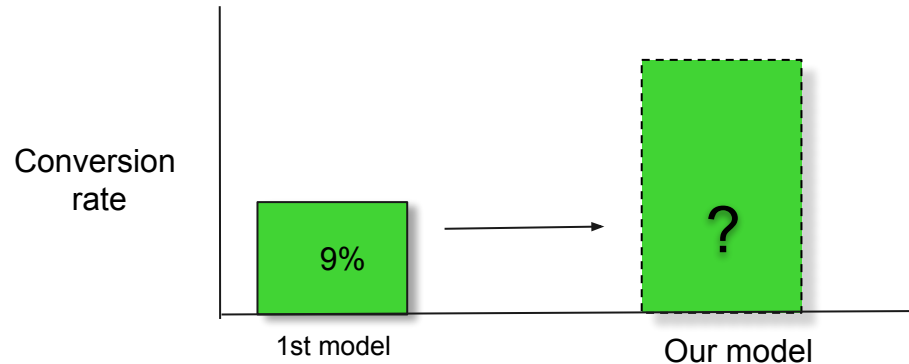
Eric Greenberg Goldy

Chengzhang (Jenny) Jiang

Tsai-Yuan (Amy) Sung

Business Understanding

- Context: who is Thera Bank?
- Problem: too many deposit (liability) customers, not enough loaners (asset) customers
- Creative solution: target current customers as opposed to attracting new ones
- Basis for a new model:



Business Goal

Determine which of Thera Bank's current deposit-only customers have a high likelihood of taking out a loan, and target them for marketing loan products.

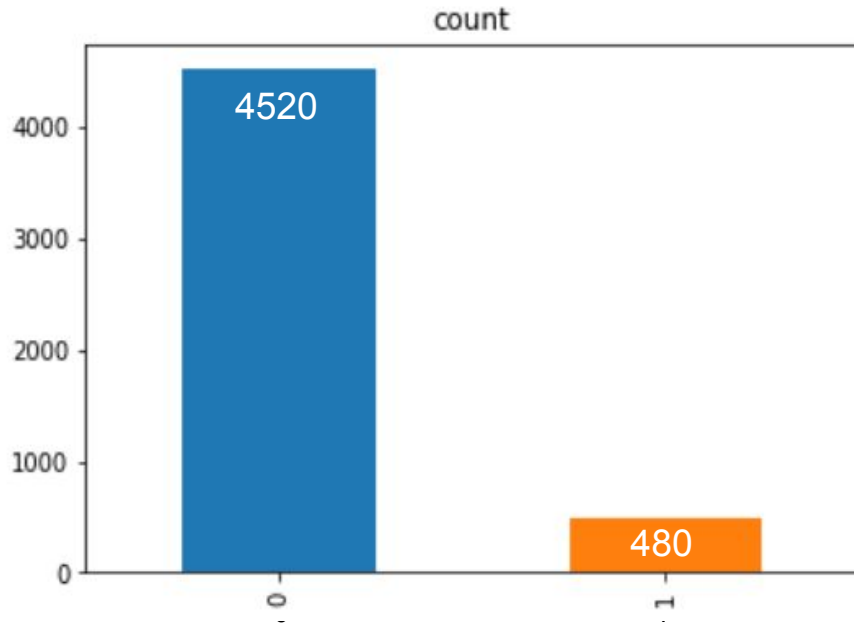


Data Understanding

	ID	Age	Experience	Income	ZIP Code	Family	CCAvg	Education	Mortgage	Securities Account	CD Account	Online	CreditCard	Personal Loan
0	1	25	1	49	91107	4	1.60	1	0	1	0	0	0	0
1	2	45	19	34	90089	3	1.50	1	0	1	0	0	0	0
2	3	39	15	11	94720	1	1.00	1	0	0	0	0	0	0
3	4	35	9	100	94112	1	2.70	2	0	0	0	0	0	0
4	5	35	8	45	91330	4	1.00	2	0	0	0	0	1	0
5	6	37	13	29	92121	4	0.40	2	155	0	0	1	0	0

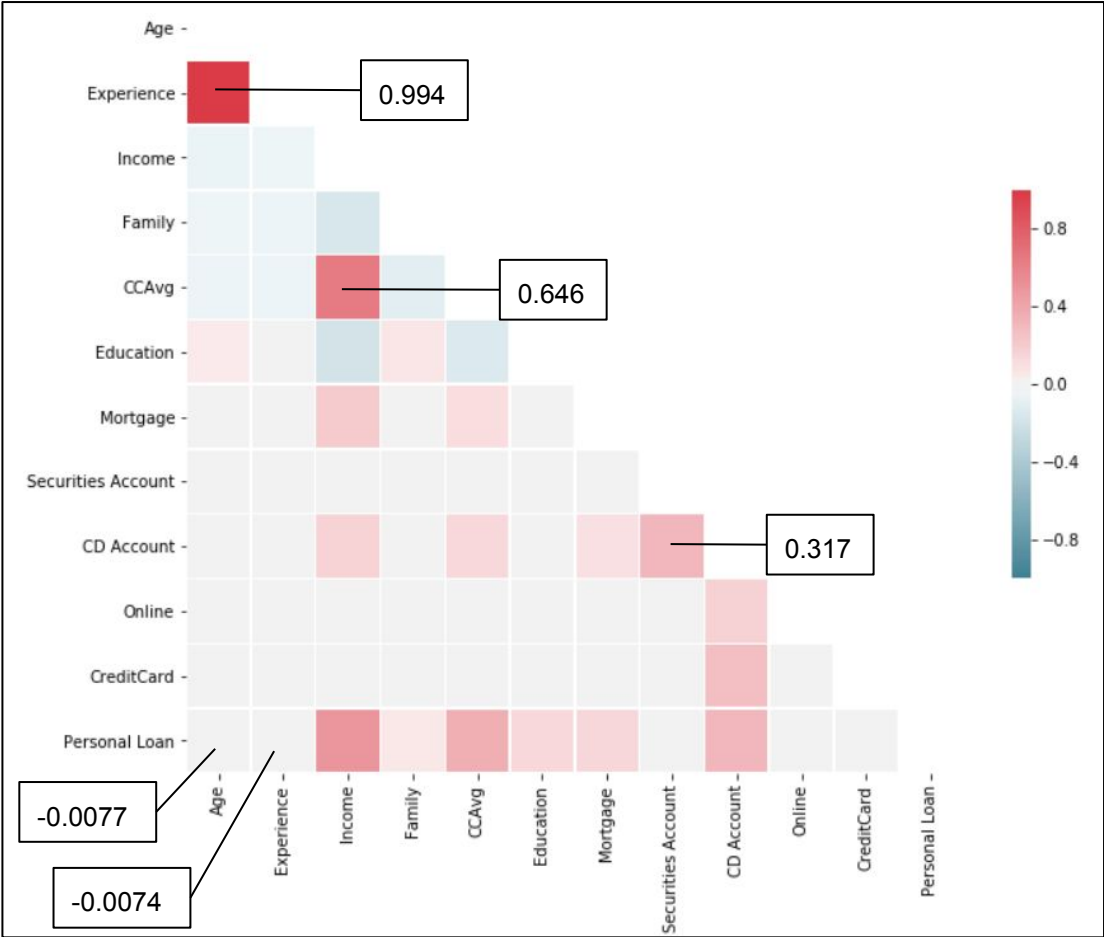
- Data Source: Kaggle
- Data Size:
5000 instances, 14 columns
- No Missing values
- Classification
- Target variable:
Personal loan

Data Preparation

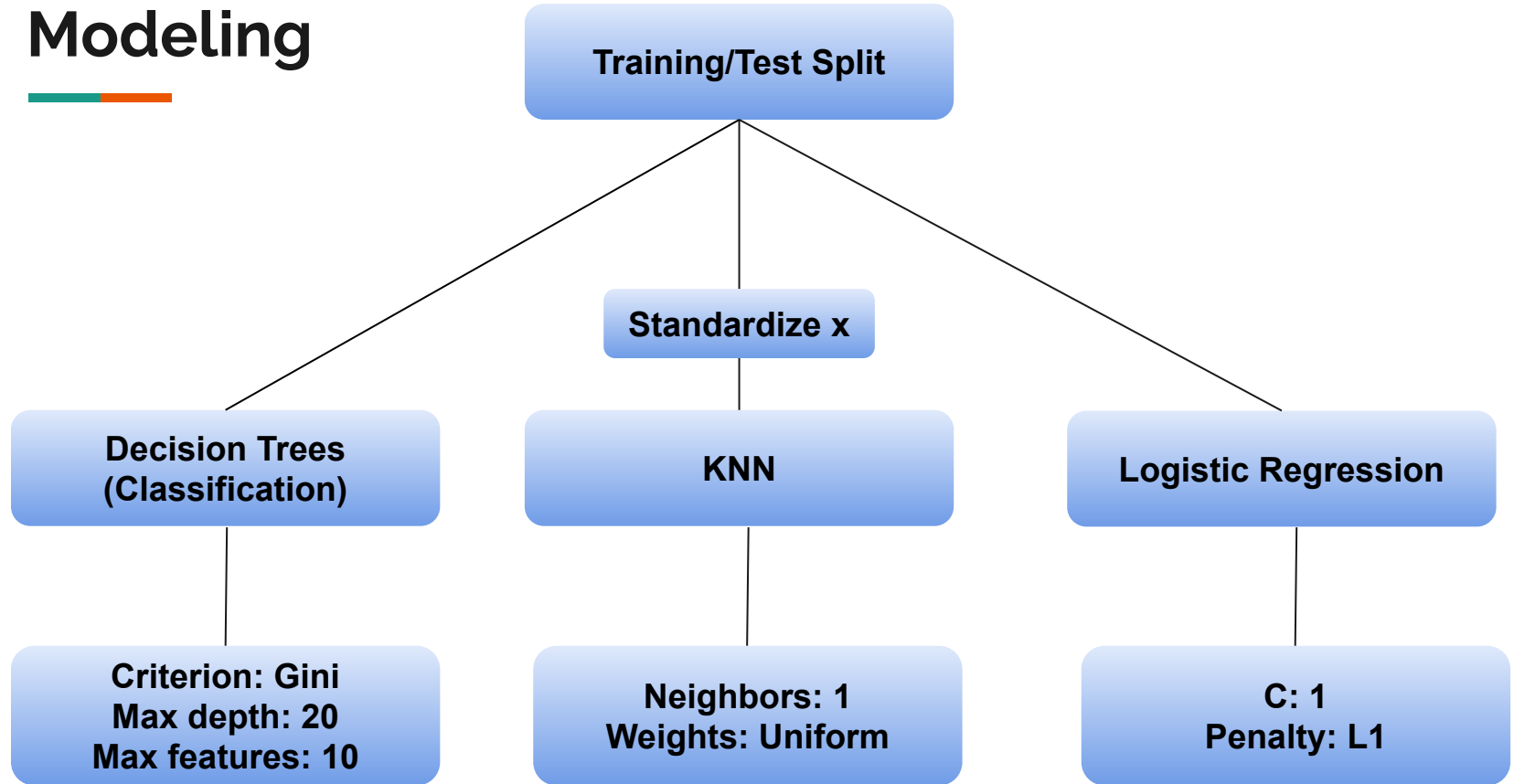


- Imbalanced Dataset - need resample
- Experience (52 Negative values)
- Category variables (Dummy variables)
 - Family (4 levels)
 - 1 - family with 1 child,
 - 2 - family with 2 children,
 - 3 - family with 3 children,
 - 4 - family with 4 children
 - Education (3 levels)
 - 1 - Undergrad,
 - 2 - Graduate,
 - 3 - Advanced/Professional

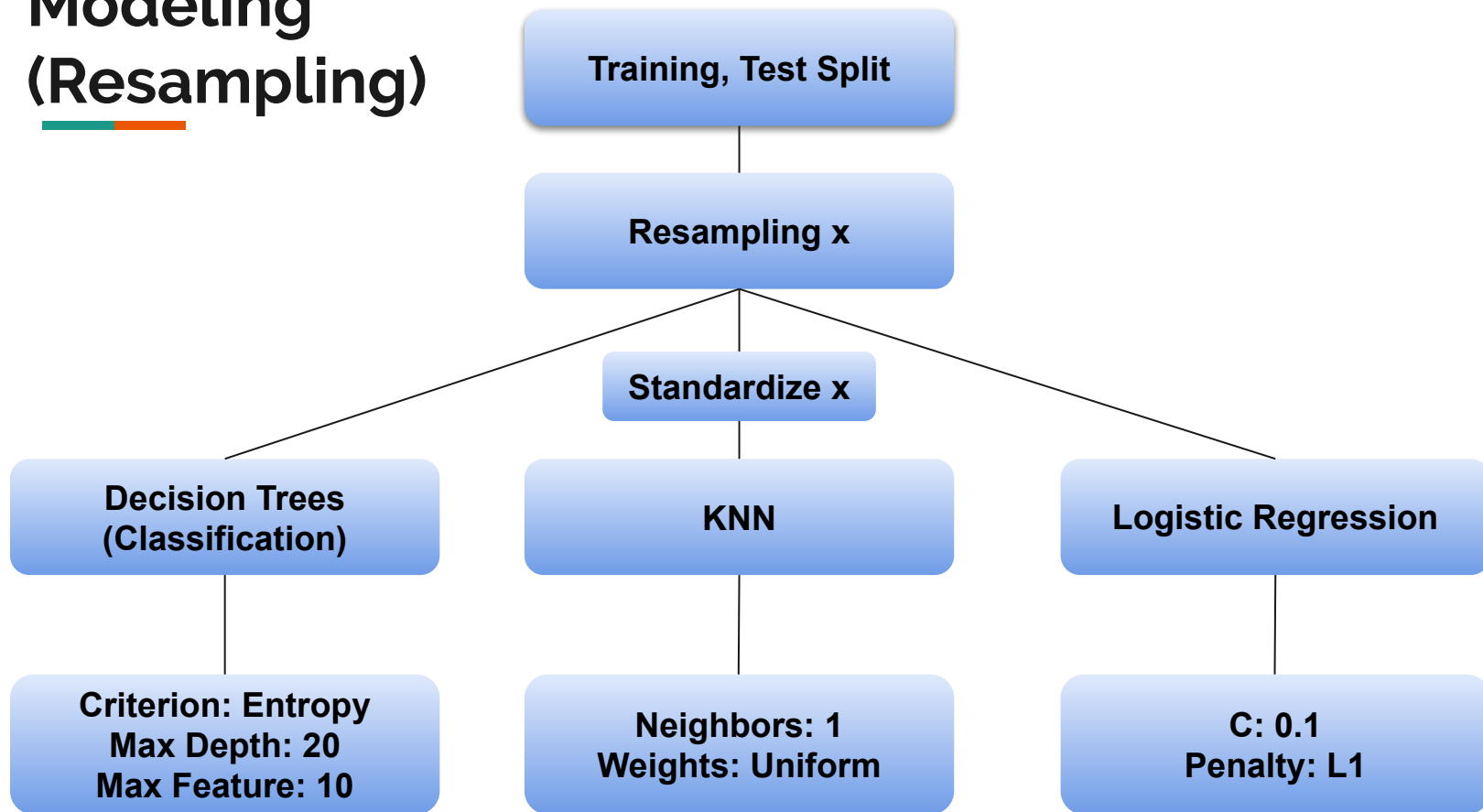
-
- Heatmap showing Pearson correlation coefficients between variables. The color scale ranges from -0.8 (blue) to 0.8 (red).
- | | Age | Experience | Income | Family | CCAvg | Education | Mortgage | Securities Account | CD Account | Online | CreditCard | Personal Loan |
|--------------------|---------|------------|---------|---------|---------|-----------|----------|--------------------|------------|---------|------------|---------------|
| Age | 1.000 | 0.994 | -0.0077 | -0.0074 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 |
| Experience | 0.994 | 1.000 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 | -0.0077 |
| Income | -0.0077 | -0.0077 | 1.000 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 |
| Family | -0.0074 | -0.0077 | 0.646 | 1.000 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 |
| CCAvg | -0.0077 | -0.0077 | 0.646 | 0.646 | 1.000 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 |
| Education | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 1.000 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 |
| Mortgage | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 0.646 | 1.000 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 |
| Securities Account | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 1.000 | 0.317 | 0.317 | 0.317 | 0.317 |
| CD Account | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.317 | 1.000 | 0.317 | 0.317 | 0.317 |
| Online | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.317 | 0.317 | 1.000 | 0.317 | 0.317 |
| CreditCard | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.317 | 0.317 | 0.317 | 1.000 | 0.317 |
| Personal Loan | -0.0077 | -0.0077 | 0.646 | 0.646 | 0.646 | 0.646 | 0.646 | 0.317 | 0.317 | 0.317 | 0.317 | 1.000 |



Modeling



Modeling (Resampling)



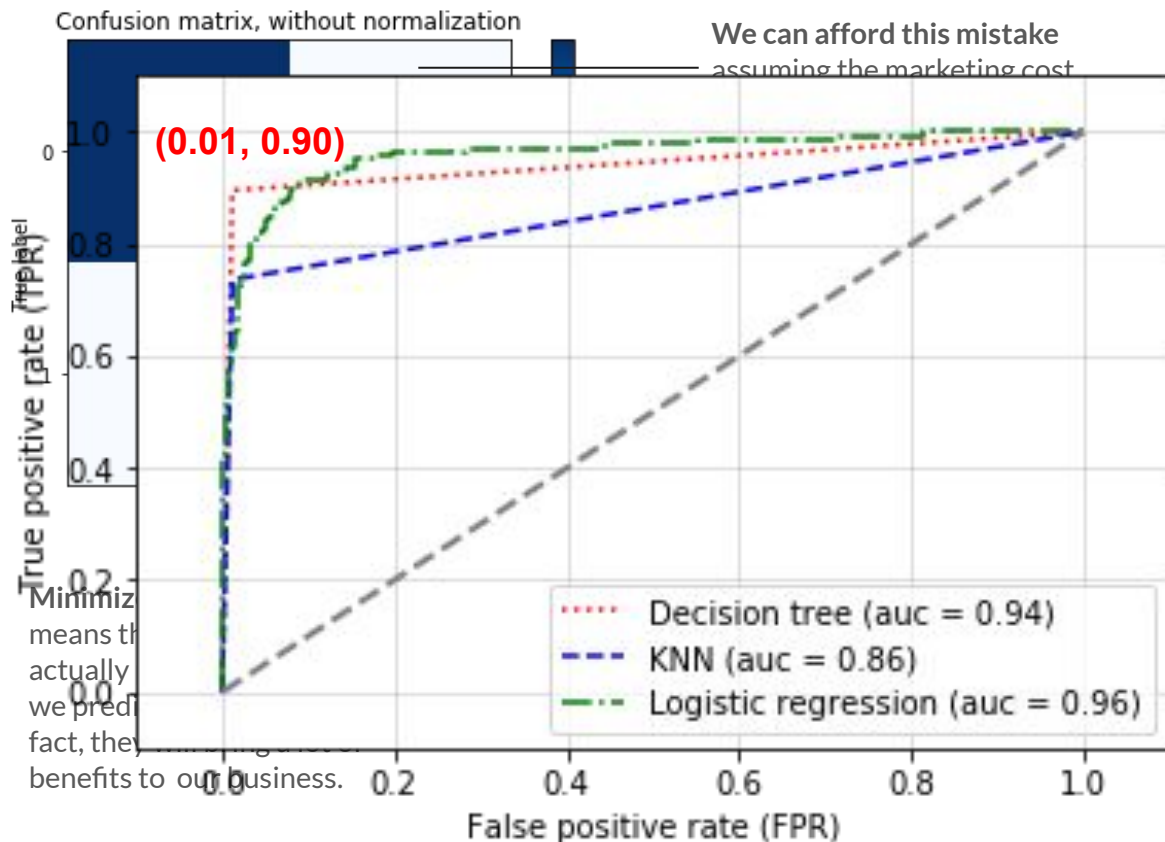
Model Performance on Test Data

model	accuracy	precision (1)	recall (1)	f-statistics (1)
Decision Tree(Before Resampling)	0.98	0.89	0.89	0.89
Decision Tree(Resampling)	0.98	0.89	0.90	0.89
KNN (Before Resampling)	0.97	0.87	0.76	0.81
KNN (Resampling)	0.96	0.87	0.74	0.80
Logistic Regression(Before Resampling)	0.96	0.88	0.65	0.74
Logistic Regression(Resampling)	0.91	0.51	0.91	0.66

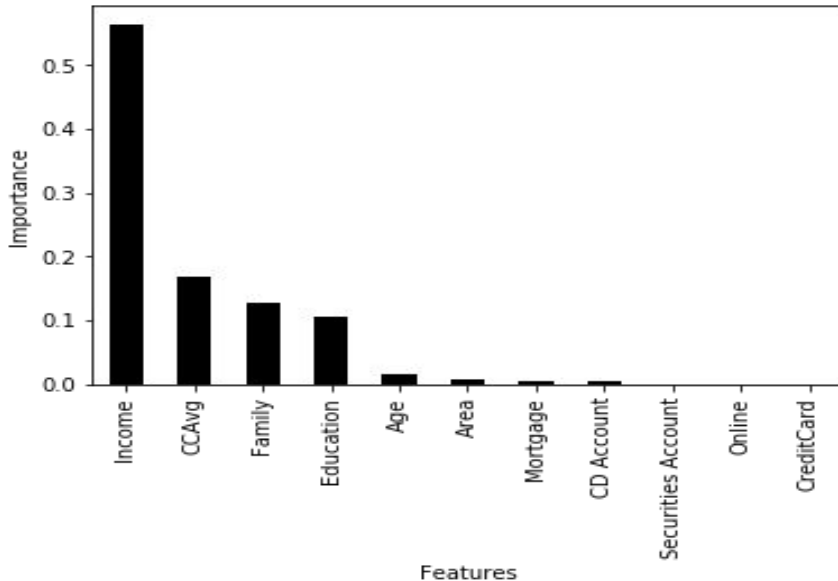
Evaluation

Our best model is the decision tree model after resampling.

- Accuracy: 0.98
- Confusion matrix
- ROC curve



Evaluation



- Business case & Expected improvement
 - Strengthen business understanding: print feature importance
 - Marketing techniques
 - Email/ Flyer attached at the end of monthly e-statement (or physical statements)
 - Phone
- ROI
 - Expect to be high
 - Low cost, high accuracy

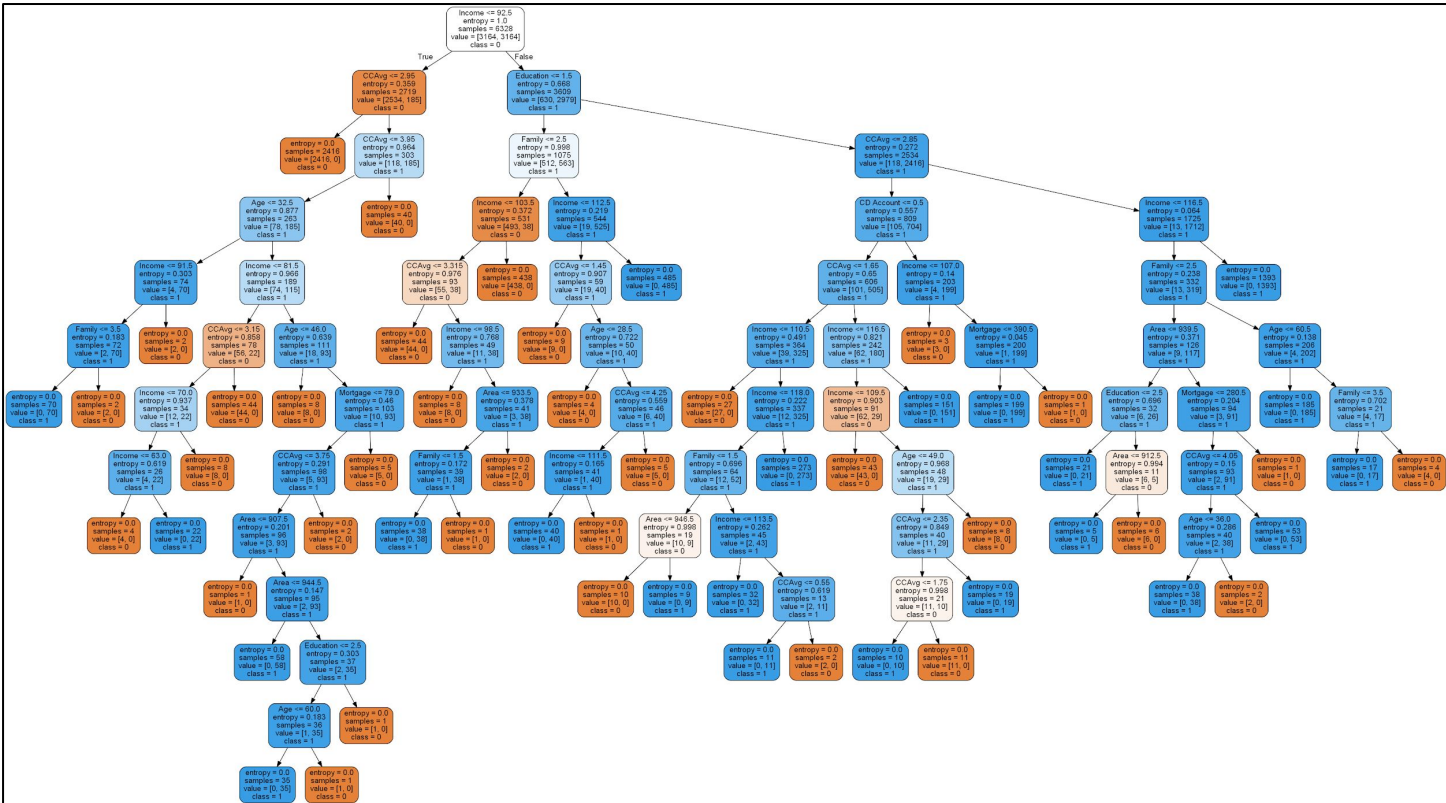
Deployment

- How will the results be deployed?
 - Targeted marketing campaigns
- Possible concerns or issues with the model:
 - Not meeting same expectations of previous model
 - Could lose customers
- Ethical problems:
 - Non-intrusive data mining
- Risks:
 - Default rate possibility:
 - We would like to also consider customer's credit scores as part of the modeling process

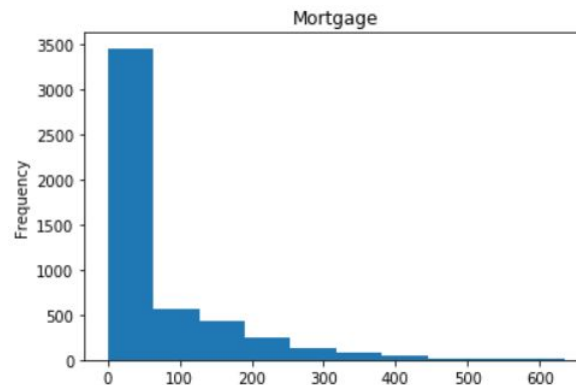
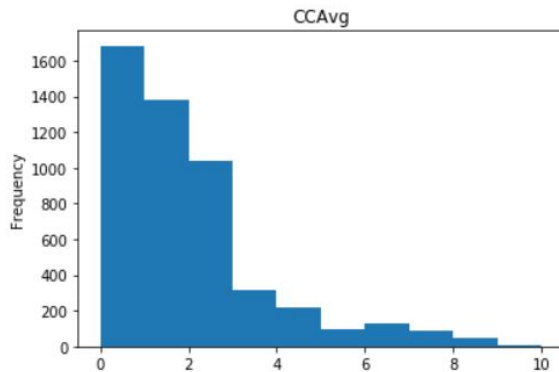
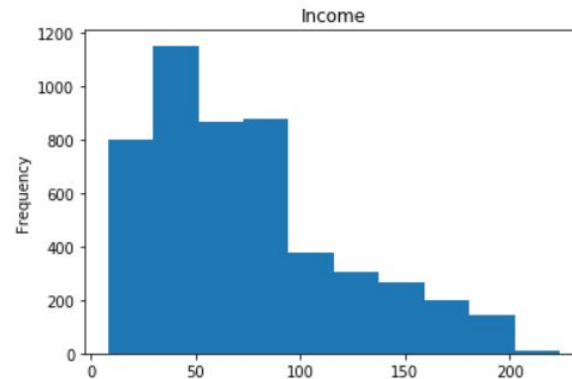
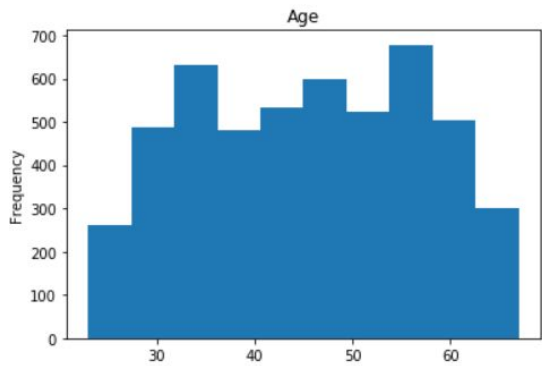




Thank You!
Any Questions?



Appendix B -- Histogram for numeric attributes




Appendix C -- Correlation Coefficients



	Age	Experience	Income	Family	CCAvg	Education	Mortgage	Securities Account	CD Account	Online	CreditCard	Personal Loan
Age	1.000000	0.994214	-0.055269	-0.046418	-0.052012	0.041334	-0.012539	-0.000436	0.008043	0.013702	0.007681	-0.007726
Experience	0.994214	1.000000	-0.046880	-0.052119	-0.050030	0.014071	-0.010950	-0.001258	0.009905	0.013765	0.009073	-0.008060
Income	-0.055269	-0.046880	1.000000	-0.157501	0.645984	-0.187524	0.206806	-0.002616	0.169738	0.014206	-0.002385	0.502462
Family	-0.046418	-0.052119	-0.157501	1.000000	-0.109275	0.064929	-0.020445	0.019994	0.014110	0.010354	0.011588	0.061367
CCAvg	-0.052012	-0.050030	0.645984	-0.109275	1.000000	-0.136124	0.109905	0.015086	0.136534	-0.003611	-0.006689	0.366889
Education	0.041334	0.014071	-0.187524	0.064929	-0.136124	1.000000	-0.033327	-0.010812	0.013934	-0.015004	-0.011014	0.136722
Mortgage	-0.012539	-0.010950	0.206806	-0.020445	0.109905	-0.033327	1.000000	-0.005411	0.089311	-0.005995	-0.007231	0.142095
Securities Account	-0.000436	-0.001258	-0.002616	0.019994	0.015086	-0.010812	-0.005411	1.000000	0.317034	0.012627	-0.015028	0.021954
CD Account	0.008043	0.009905	0.169738	0.014110	0.136534	0.013934	0.089311	0.317034	1.000000	0.175880	0.278644	0.316355
Online	0.013702	0.013765	0.014206	0.010354	-0.003611	-0.015004	-0.005995	0.012627	0.175880	1.000000	0.004210	0.006278
CreditCard	0.007681	0.009073	-0.002385	0.011588	-0.006689	-0.011014	-0.007231	-0.015028	0.278644	0.004210	1.000000	0.002802
Personal Loan	-0.007726	-0.008060	0.502462	0.061367	0.366889	0.136722	0.142095	0.021954	0.316355	0.006278	0.002802	1.000000

Appendix D -- Data Types



ID	int64
Age	int64
Experience	float64
Income	int64
ZIP Code	category
Family	category
CCAvg	float64
Education	category
Mortgage	int64
Personal Loan	int64
Securities Account	int64
CD Account	int64
Online	int64
CreditCard	int64
Experience Median	float64
Area	category
dtype: object	

Appendix E -- Best Parameters (Pre-resampling)

```
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=20,  
    max_features=10, max_leaf_nodes=None,  
    min_impurity_decrease=0.0, min_impurity_split=None,  
    min_samples_leaf=1, min_samples_split=3,  
    min_weight_fraction_leaf=0.0, presort=False, random_state=0,  
    splitter='best')
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',  
    metric_params=None, n_jobs=1, n_neighbors=1, p=2,  
    weights='uniform')
```

```
LogisticRegression(C=1, class_weight=None, dual=False, fit_intercept=True,  
    intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,  
    penalty='l1', random_state=0, solver='liblinear', tol=0.0001,  
    verbose=0, warm_start=False)
```

Appendix F -- Best Parameters (Resampling)



```
DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=20,  
    max_features=10, max_leaf_nodes=None,  
    min_impurity_decrease=0.0, min_impurity_split=None,  
    min_samples_leaf=1, min_samples_split=2,  
    min_weight_fraction_leaf=0.0, presort=False, random_state=0,  
    splitter='best')
```

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',  
    metric_params=None, n_jobs=1, n_neighbors=1, p=2,  
    weights='uniform')
```

```
LogisticRegression(C=0.1, class_weight=None, dual=False, fit_intercept=True,  
    intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,  
    penalty='l1', random_state=0, solver='liblinear', tol=0.0001,  
    verbose=0, warm_start=False)
```