

Submitted by: Saurav Premkumar, sp7818

Business Case

The dataset holds value for a variety of organizations that require an assessment of an individual's financial background. For banks, specifically, this information is key to predicting whether a customer might accept a personal loan based on various indicators. When this dataset is utilized in targeted marketing efforts, it offers several benefits. It allows for more effective marketing campaigns by identifying the right customer demographics, improves the likelihood of loan acceptance by reaching out to those with a higher probability to accept, ensures better use of marketing resources by targeting likely candidates, and increases customer contentment by offering relevant loan options that align with their financial profiles.

Q1. Data Pre-processing

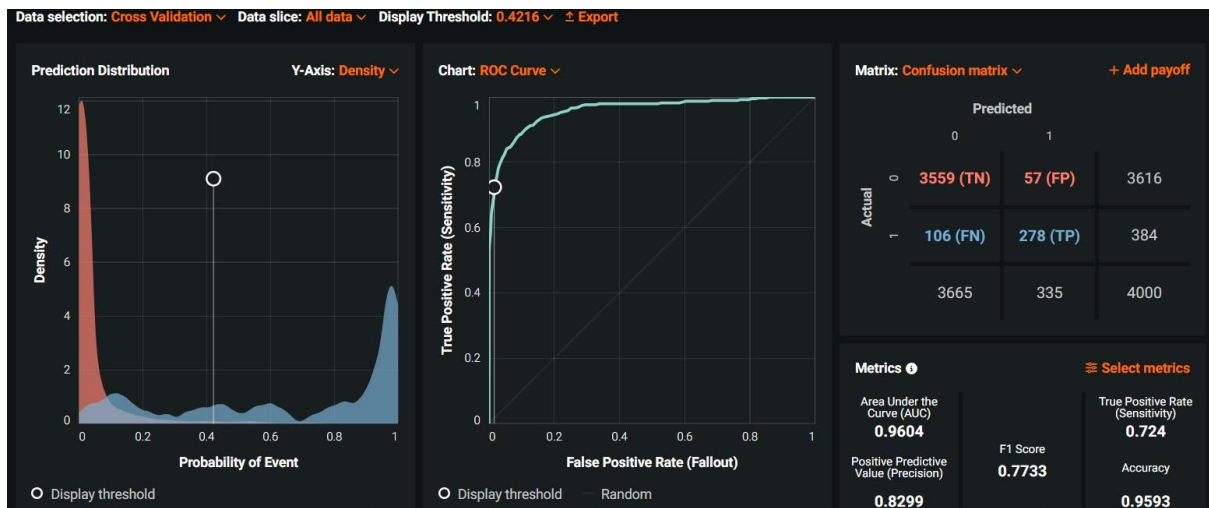
- Selected the following features for modelling: Personal Loan, Income, Mortgage, Online, Credit card, CCAvg, CD Account, Securities Account, Education, Family
- Recorded Education and Family as categorical.

<input type="checkbox"/>	CCAvg	7	<div><div></div></div>	Numeric	104	0	1.94	1.75	1.50	0	10
<input type="checkbox"/>	CD Account	12	<div><div></div></div>	Numeric	2	0	0.06	0.24	0	0	1
<input type="checkbox"/>	CreditCard	14	<div><div></div></div>	Numeric	2	0	0.30	0.46	0	0	1
<input type="checkbox"/>	Income	4	<div><div></div></div>	Numeric	162	0	74.08	46.34	64	8	224
<input type="checkbox"/>	Mortgage	9	<div><div></div></div>	Numeric	319	0	55.77	102	0	0	635
<input type="checkbox"/>	Online	13	<div><div></div></div>	Numeric	2	0	0.60	0.49	1	0	1
<input type="checkbox"/>	Personal Loan	10	Target	Numeric	2	0	0.10	0.29	0	0	1
<input type="checkbox"/>	Securities Account	11	<div><div></div></div>	Numeric	2	0	0.10	0.30	0	0	1
<input type="checkbox"/>	Updated_Education_Categorical	8	<div><div></div></div>	Categorical	3	0					
<input type="checkbox"/>	Updated_Family_Categorical	6	<div><div></div></div>	Categorical	4	0					

Q2. Model Performance Evaluation

Recall	0.724
Precision	0.829
F1	0.773
Accuracy	0.959
Error rate	0.041
ROC AUC	0.96

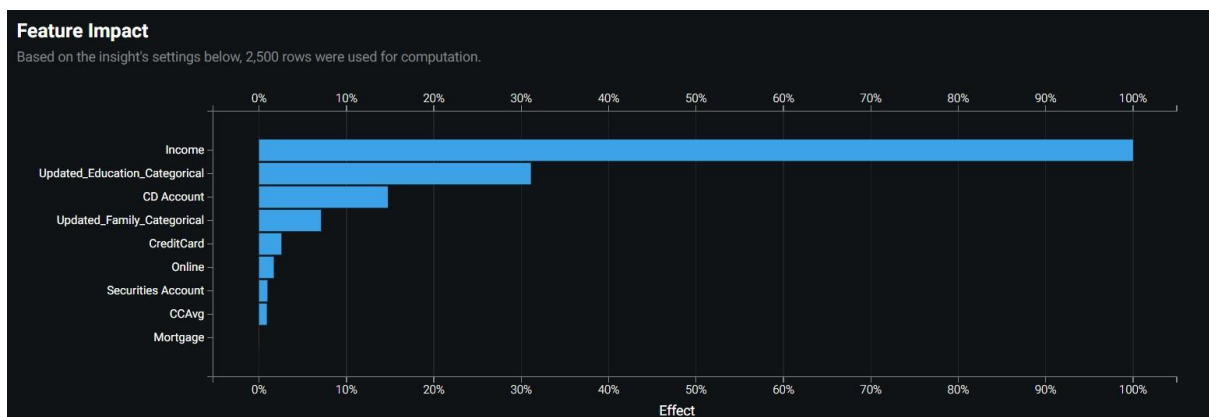
Model metrics are reported at a threshold value of 0.4216 which optimized the F1 value at 0.773.



The accuracy of the naïve model that predicts the majority class (not accepting the personal loan) is approximately 90.4%. Logistic regression outperforms the naïve model.

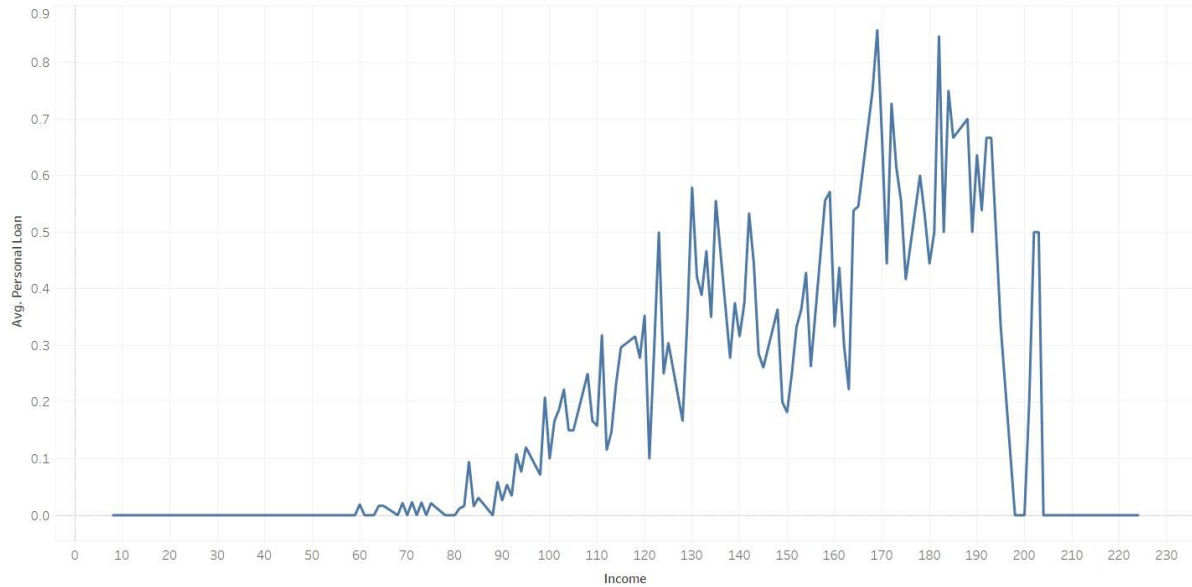
Q3. Feature Impact

Income, Education, CD Account, Family, Credit card, Online, Securities Account, CCAvg had a significant effect on the model.



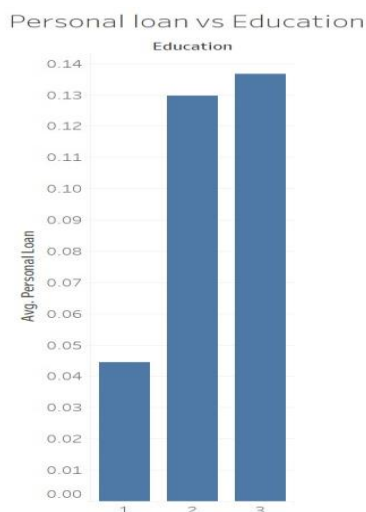
Personal Loan vs Income:

Personal loan vs Income



The line graph suggests that the average acceptance rate for personal loans increases with income, especially beyond an income level of approximately 120k dollars, indicating a potential positive correlation between income and the likelihood of accepting personal loans.

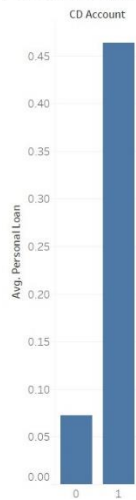
Personal Loan vs Education:



The bar chart shows that customers with higher education levels (presumably labeled 2 and 3) have a similar and notably higher average acceptance rate for personal loans compared to those with the lowest education level (labeled 1)

Personal loan vs CD Account:

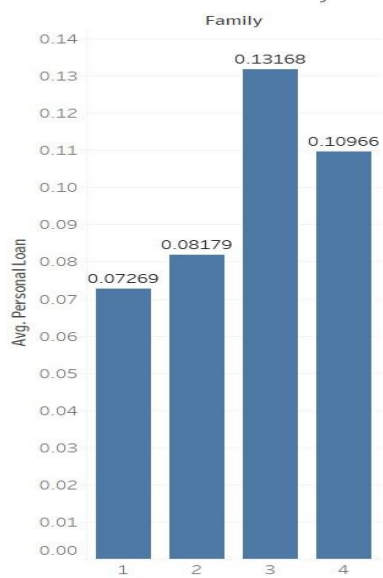
Personal loan vs CD Account



The bar chart indicates that customers with a certificate of deposit (CD) account (labeled 1) have a substantially higher average acceptance rate for personal loans than those without a CD account (labeled 0).

Personal loan vs Family:

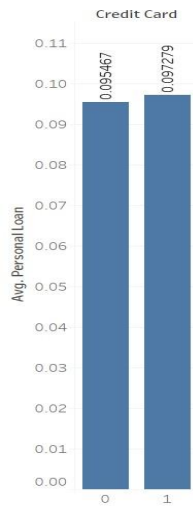
Personal loan vs Family



The bar chart reveals that customers with a family size of 3 have the highest average acceptance rate for personal loans, followed by those with a family size of 4, while those with family sizes of 1 or 2 have lower acceptance rates.

Personal loan vs Credit Card:

Personal loan vs Credit Card



The bar chart shows that customers without a credit card (labeled 0) have a slightly higher average acceptance rate for personal loans than those with a credit card (labeled 1).

Q4. Model Sanity Check:

The model has predictive capability. ROC AUC= 0.96 F1 at optimal threshold is 0.773.

The key features imply that the predictive outcome makes sense. It is likely that more personal loans were given to people with above average incomes, with higher education, and with CD account. Another thing, that should have impact was the credit card, but it had comparatively lesser impact on the probability of personal loan.