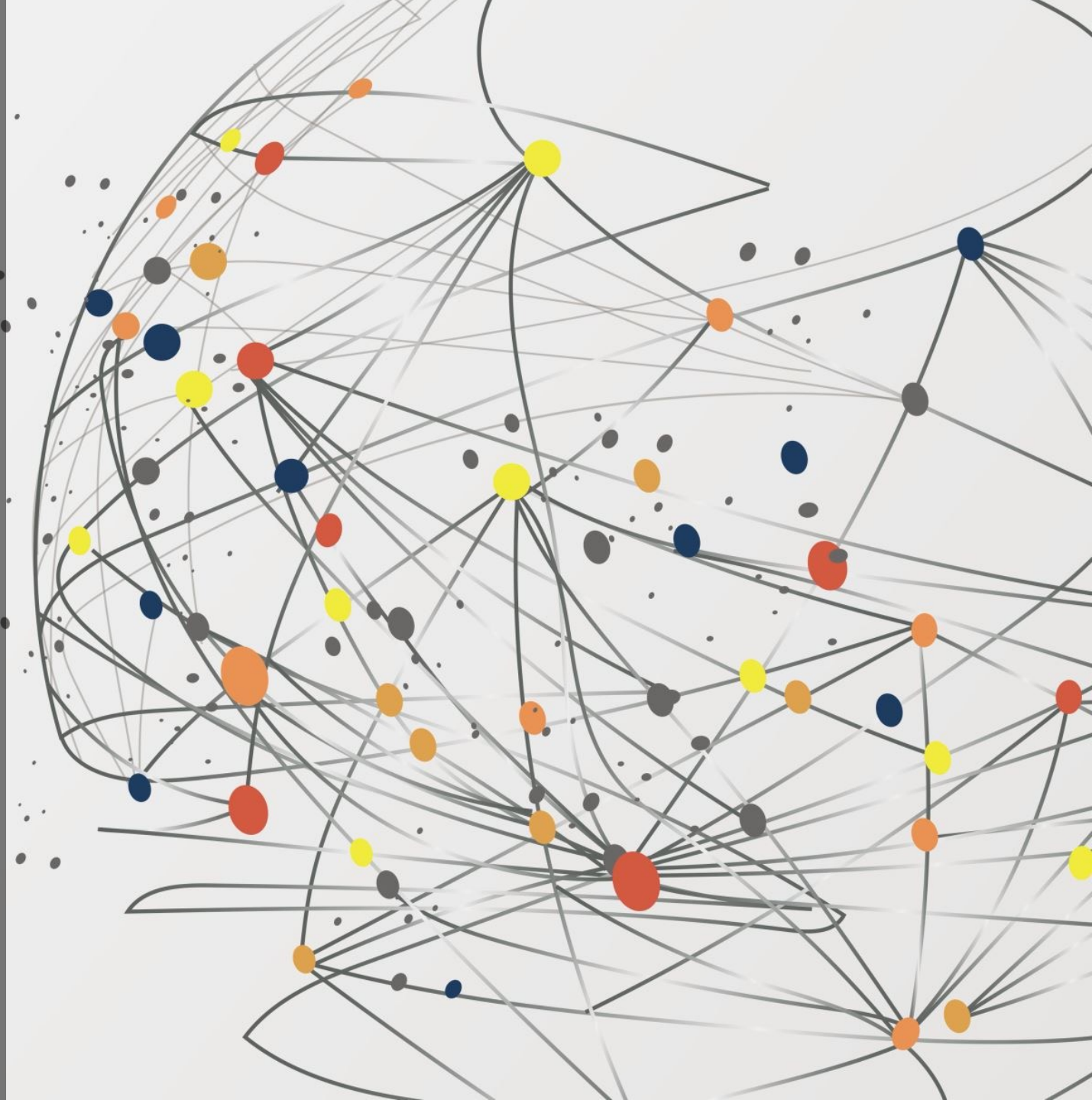
Prediction System – Customer Loan Acceptance

-Jyothi H



Business Questions

How do we accurately build a personal loan acceptance prediction system for a liability customer?

Key Questions To Address

The following questions need to be answered to effectively predict if a liability customer will buy the loan or not

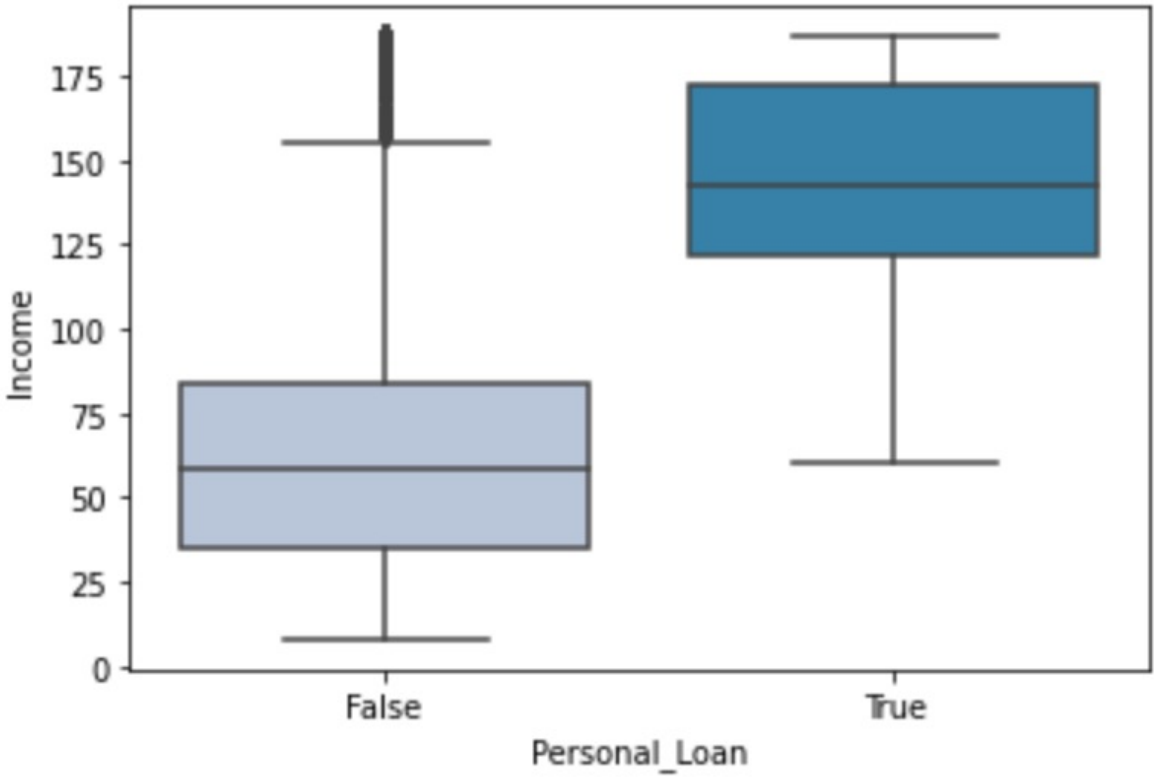
- Does various predicting factors really affect the likelihood of customer loan acceptance?
- What are the importance variables affecting the likelihood of customer loan acceptance?
- Does the income of the customer really effect the likelihood of customer loan acceptance?
- Does the family size of the customer really effect the likelihood of customer loan acceptance?
- Does customer education really effect the likelihood of customer loan acceptance?
- Do age and professional experience really effect the likelihood of customer loan acceptance?
- Does customer's existing relationship with the bank(various account holdings) really effect the likelihood of customer loan acceptance?

Classification Techniques & Models Used For Prediction

- Logistic Regression Model
  - Default Logistic Regression using sklearn
  - Logistic Regression with AUC-ROC Curve Threshold
  - Logistic Regression with Precision-Recall Curve Threshold

- Decision Tree Model
  - Default Decision Tree with No Pruning
  - Decision Tree with Maximum Depth Restriction
  - Decision Tree with Hyperparameter Grid Search
  - Decision Tree with Cost Complexity Pruning

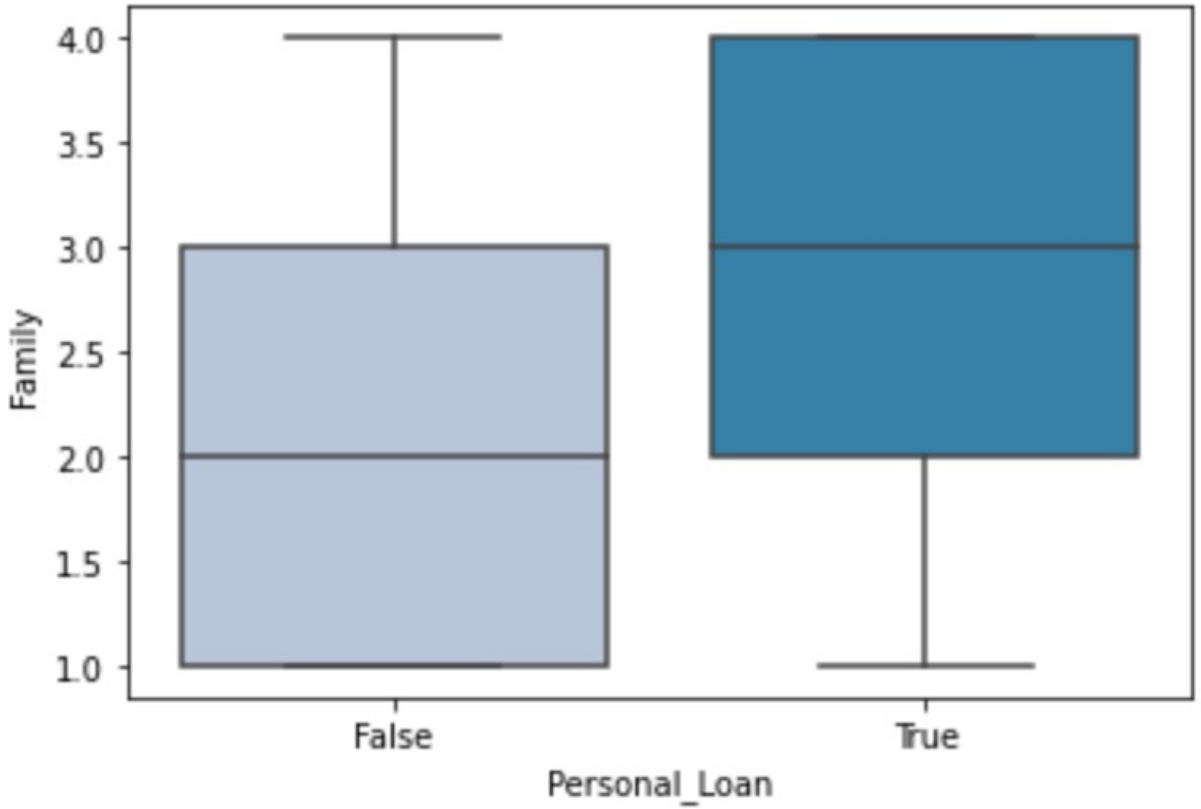
Correlation – Personal Loan Vs Income



As depicted in the graph, the likelihood of a customer buying a loan tends to be positively correlated with the Income of a customer.

Therefore, we would expect to see this as an important feature in our model prediction.

Correlation – Personal Loan Vs Family Size



The graph shows that a larger family size is associated with positive correlation with the likelihood of loan purchase by the customer.

Therefore, we should expect to see Family as an important feature in our model prediction.

Correlation (EDA)– Account Related Variables (Categorical)

Features	Did Not Accept Past Loan	Did Accept Past Loan	Significance
Has a Security Account ?	10%	10%	It appears that Possessing a Security Account with the bank does not seem to have any bearing on weather a customer will purchase the bank loan or not
Has a CD Account ?	4%	30%	Since 30% of the customers who has a CD account has accepted the loan in the past marketing campaign, as opposed to 4% who did not accept , owning a CD account might have a bearing on the loan acceptance likelihood .
Use Online Banking ?	60%	60%	Since customers who use online banking are just as likely to accept the loan as they are to not accept, this parameter does not seem to have any bearing on loan acceptance
Posses Credit Card from different bank ?	30%	30%	Since customers who use posses credit card from another bank are just as likely to accept the loan as they are to not accept, this parameter does not seem to have any bearing on loan acceptance

Model Evaluation Criteria

Possible Errors in Prediction:

1. Error 1 - Predicting a liability customer will buy a personal loan but in reality, does not.
    - Leads to misuse of marketing expenditure and resources
2. Error 2 - Predicting a liability customer will not buy a personal loan but in reality, he will –
    - Results in potential lost opportunity

Which Case Holds More Weight ?

- Error 1 results in misuse of capital which can be critical for small cash strapped companies
- Error 2 results in lost opportunity, which is important for stable companies with healthy balance sheet, looking to grow.

We will assume both errors to be equally important and therefore will try to maximize the F1_Score to get an optimal model

Model Prediction - Feature Importance

Features	Logistic Regression (Coefficient)	Decision Tree (Importance)
Income	0.059085	0.330314
Family	0.578179	0.266687
Education (Graduate)	3.785787	0.174482
Education (Professional)	3.811852	0.140037
Credit Card Spending	0.153516	0.044067
Has CD Account	3.903719	0.025605

Model Prediction – F1 Score for Best Model

Features	**Logistic Regression** (Precision-Recall Curve Threshold of 0.32)	**Decision Tree** (Cost Complexity Pruning)
Training Data	0.77	0.95
Testing Data	0.78	0.90

Best Model Recommendation : Decision Tree with Cost Complex Pruning

Potential Benefits Of Accurate Prediction Model

- Increased Marketing ROI – By accurately targeting the right segment of customers, the marketing team can maximize its ROI on marketing spend.

- Increased Conversion Rate – Accurate prediction model allows the company to target the right customers , thereby increasing the operational KPI's of the company , mainly the conversion wrt loan acceptance.

- Improved Profit Margin – Optimal use of resource and capital that is targeted towards intended customers helps company in improving the profit margin.

Recommendation

The decision model with post-pruning results in a F1_score of 0.90 for training data set and 0.95 for testing data set, given a set of features that are described by the regression model. Therefore, I recommend leveraging this model using the below mentioned key factors , along with their corresponding coefficients described earlier) to determine the likelihood of customer accepting a loan

- Customer Income
- Customer Family Size
- Customer Education
- Customer Credit Card Spending
- Customer CD Account with the Bank

In conclusion, I recommend the marketing team to target a customer segment with higher income, larger family size, with higher level of education , and higher monthly credit card spending , who has a CD account with the bank for its campaigning.