Data Science Team

Potential loan-taker prediction modeling project

Business Report

Student ID: s3741327
Student name: Juyeon Kim
27/05/2022

TABLE OF CONTENTS

- Understanding business problems & objectives
- Insights from historical data
- Machine learning model evaluation
- Further recommendations

Accuracy is a good measure when the target variable class in the data are nearly balanced.

Business Understanding

Business issues

- ***Not successful campaign result*** last year: only 16.6% customers took a loan from campaign
- ***Cost ineffective*** campaign: was designed to all range of customers

Business goal

Want to launch ***a targeted campaign*** to only to ***potential customers*** who are likely to take a loan

Business Understanding -2

Machine learning prediction model

Step 1. Understand the historical data
Step 2. Data cleaning
Step 3. Modeling
Step 4. Tuning to increase model accuracy

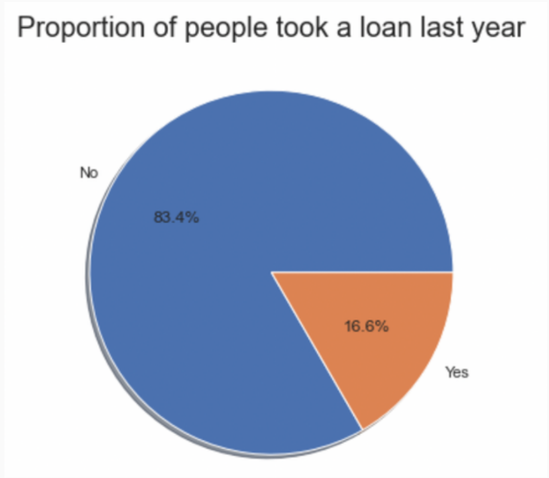
Expected outcome

(1) Cost savings.
- only perform marketing to targeted customers

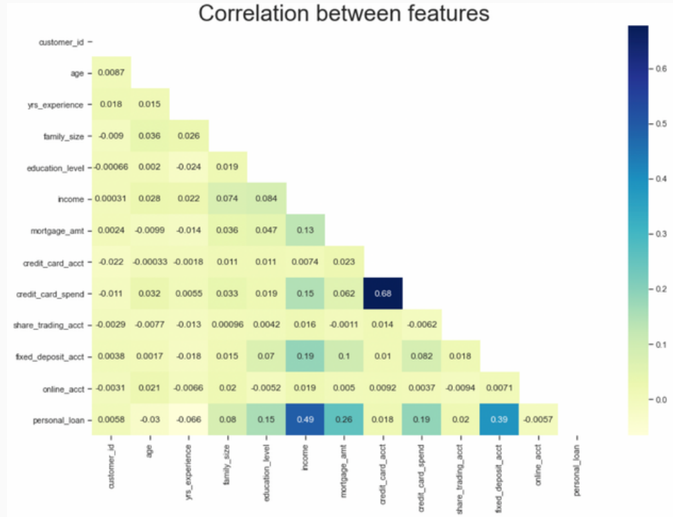
(***2) Further business opportunities***
- will bring business to be tech-savvy industry leader

Insights from historical data

Correlation Analysis



	personal_loan
personal_loan	1.000000
income	0.491728
fixed_deposit_acct	0.388889
mortgage_amt	0.259120
credit_card_spend	0.188654
education_level	0.154946
family_size	0.079590
yrs_experience	-0.065926
age	-0.029942
share_trading_acct	0.019911
credit_card_acct	0.017924
customer_id	0.005838
online_acct	-0.005674



› Interpretation

- **Last year campaign** was only worked out for **16.6%** of customers
- There is still **83.4%** of customer where we can take advantage of targeted approach

› Interpretation

Highly correlated variables to personal loan
(1) income (**49%**)
(2) fixed deposit account (**39%**)
(3) and mortgage amount (**26%**)

Highly correlated variable sets
(1) credit card spend & credit cart account (**68%**)

Insights from historical data 2

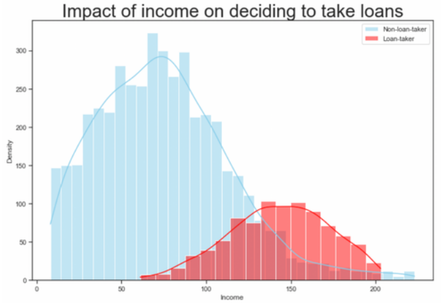
Correlation Visualization

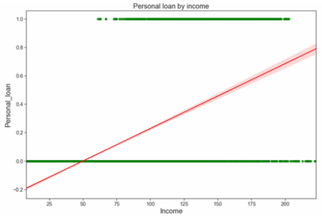
1. ***Loan and Income***

- The more income, the more likely to take a loan.

- Many of non-loan taker took a loan when their income is around $60k.

- Income of loan-taker is higher than the income of non-loan-taker.

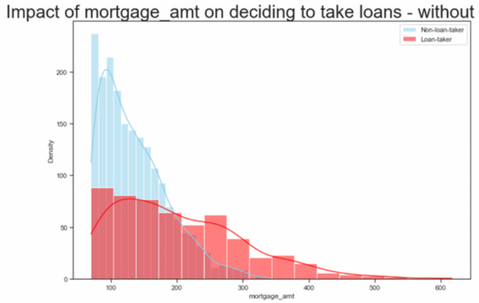


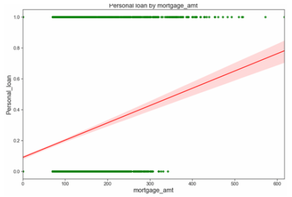


2. Loan and mortgage amount

- The more mortgage amount is, the more likely to take a loan

- The number of loan taker reached peak at the mortgate amount around $100k.

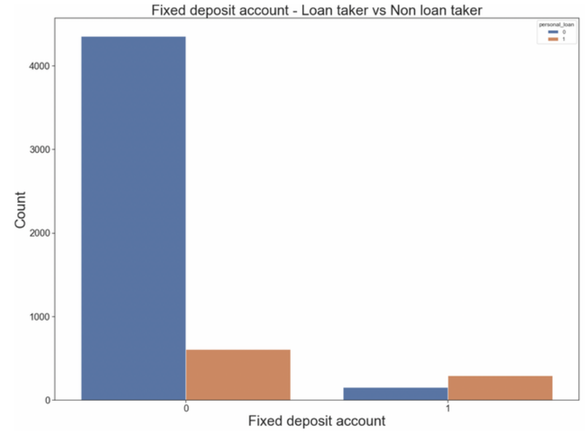




3. Loan and Fixed deposit account

- People with fixed deposit account are more likely to take a loan.

- Majority of customers doesn't have a fixed deposit account.



Model Evaluation & Best Model Selection

Random Forest vs Decision tree model

1

Accuracy Score

Random Forest model can increase predict the potential loan taker up to 90.33% after tuning.

Accuracy score	KNN	Random Forest
Before feature selection		
Before tuning	0.8816666666666667	0.8925
After tuning	0.8841666666666667	0.8941666666666667
After feature selection		
Before tuning	0.8983333333333333	0.8991666666666667
After tuning	0.8933333333333333	0.9033333333333333

2

Precision

Random Forest model can precisely predict the potential loan taker 90 people out of 100.

Weighted average of Precision	KNN	Random Forest
Before feature selection	0.87	0.86
After feature selection	0.89	0.90

3

Recall

Random Forest model can better predict 90 people as potential taker out of 100 actual number of loan taker.

Weighted average of Recall	KNN	Random Forest
Before feature selection	0.88	0.89
After feature selection	0.89	0.90

3

AUC

Higher AUC = Better predict yes/or. Random Forest model has the highest AUC of 0.77.

AUC	KNN	Random Forest
Before feature selection	0.72	0.72
After feature selection	0.75	0.77

"Tuned Random Forest model with **9 features"**
is chosen for loan prediction model.

['age', 'yrs_experience', 'family_size', 'education_level', 'income', 'mortgage_amt', 'credit_card_spend', 'share_trading_acct', 'fixed_deposit_acct']

Recommendations to management

- ***Data Collection phase***

1. The **income, mortgate_amt, fixed_deposit_acct** would be critical to model performance. **Try to avoid having missing data** on those features.
2. Try to **integrate invaluable customer data across business units**. Keep in mind that the more dataset, the more accurate the model will be.
3. **Avoid data siloed culture**. Promote data-driven culture and active data-sharing by transforming divisional view into holistic approach.

- ***Implementation phase***

4. **Value** of model implementation **must be measured and visible** with **quantifiable metrics**. Unless business leader see the value of the model, there will be no use of it.

- ***Further improvement***

5. To ensure dataset is valid, **further technique for data balancing** is required every time new features are added.

Thanks for watching

Juyeon Kim
s3741327