PROFILING CUSTOMERS THROUGH CREDIT RISK ASSESSMENT

Team Ice Kacang

BANK ISLAM

Our Team

Ng Qian Ying

Yap Jing Bo

Tan Kay Li

Chun Hanssen

Executive summary

Theme Profiling customers through credit risk assessment

Dataset Survey of Consumer Finances 2010, 2013, 2019 (US)

Questions

- How can banks understand its consumer market better?
- How can banks improve customer engagement and relationship by providing relevant products based on its customer's profile?
- How can banks respond better to each and every customer based on their credit risks?

Credit risk remain as a longstanding method for lenders to manage their credit risk

because if customers don't repay their credit, the lender loses money

S&P Dow Jones Indices and Experian:

- Bank card default rate (+2.27%)
- Auto loan default rate (0.56%)
- Mortgage default rate (0.35%)

(Source: S&P Dow Jones Indices and Experian, 2022)

Customer profiling process

Process flow for customer profiling

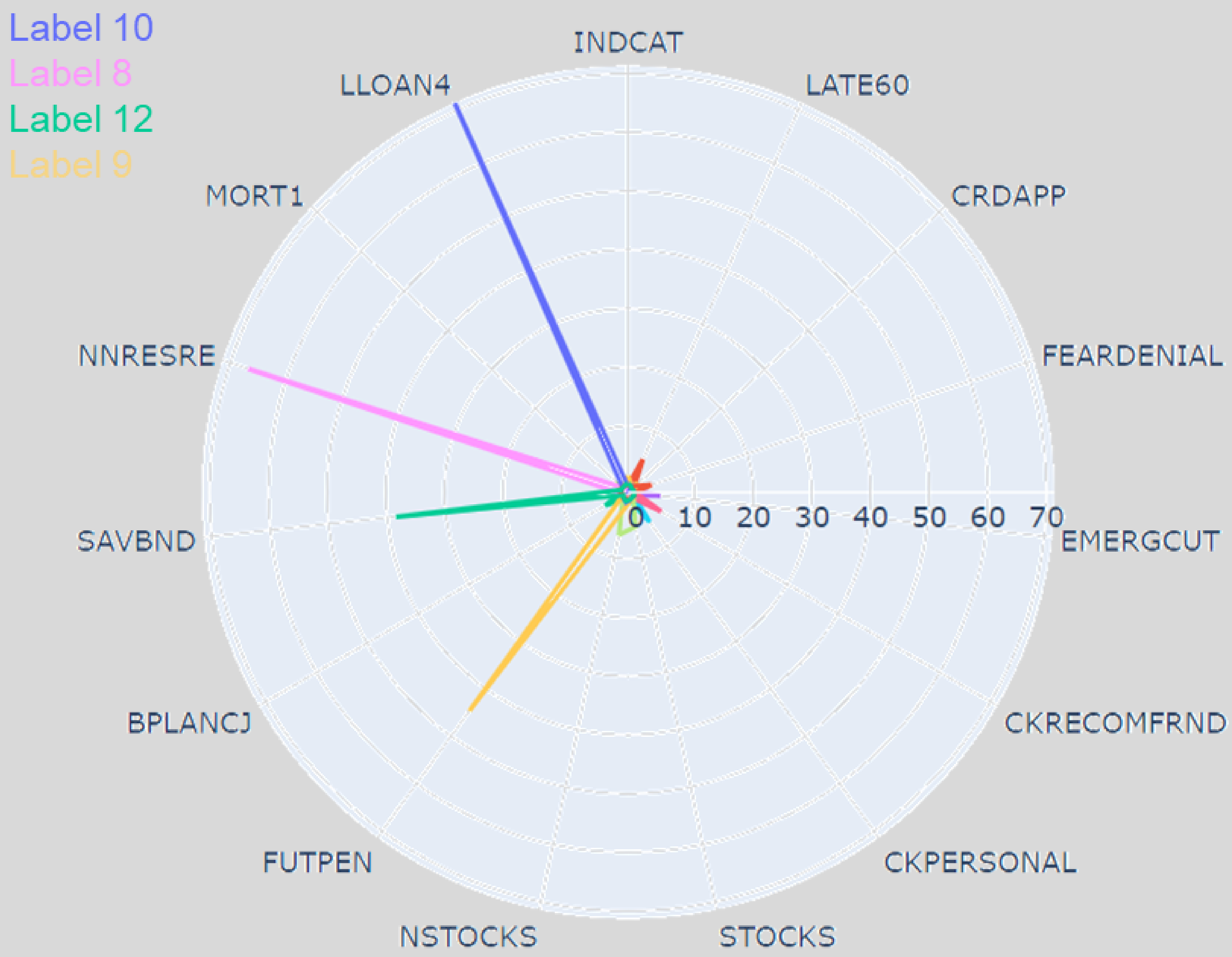
Customer is the process of grouping **similar customers** into groups with **similar characteristics**

Data → Descriptive analytics → Diagnostic analytics (Cluster analysis) → Predictive analytics

Data	Descriptive analytics	Diagnostic analytics (Cluster analysis)	Predictive analytics
Obtain sufficient data with depth for analysis	Understand about the general characteristics	Segment customers into categories	Subsequent analysis to predict the class of new customer

Cluster analysis

13 customer groups were generated through using K-means clustering



Category	Most distinct feature
Label 10	LLOAN4
Label 8	NNRESRE
Label 12	SAVBND
Label 9	FUTPEN

- LLOAN4: *Total balance of household loans where the lender is finance, loan or leasing company, or inc debt consolidator, 2019 dollars*
- NNRESRE: *Total value of net equity in nonresidential real estate held by household, 2019 dollars*
- SAVBND: *Total value of savings bonds held by household, 2019 dollars*
- FUTPEN: Future pensions (accumulated in an account for the R/S), 2019 dollars

Customer profile based on credit risk

Analysis of 2 clusters out of 13 clusters generated

Characteristics

Cluster 0

- Low proportion of high risk consumer
- Educated (avg. in college level and above)
- Low debt to income ratio in average

Cluster 10

- Contains the highest proportion of high risk consumer
- Educated (avg. in college level and above)
- High total balance of household loans
- Mostly married

How do we encourage people to retain a higher credit score?

To give benefits to high-credit-score customers as to inspire everyone to attain a higher credit score while ensuring self benefits.

- Increase the interest rate for credit card holders and loan borrowers if their credit score is low.
- Annual appreciation for those high-credit-score customers such as a gift, vouchers, etc.
- Other current benefits.

Potential value

Through enhancing customer credit risk profile, banks could...

Strategize offers and marketing strategies based on customer's value

Provide assistance or alternatives to high risk customers

Ensure stable money flow in the bank

Thank You

Appendix

Methodology

Data cleaning and modelling techniques applied

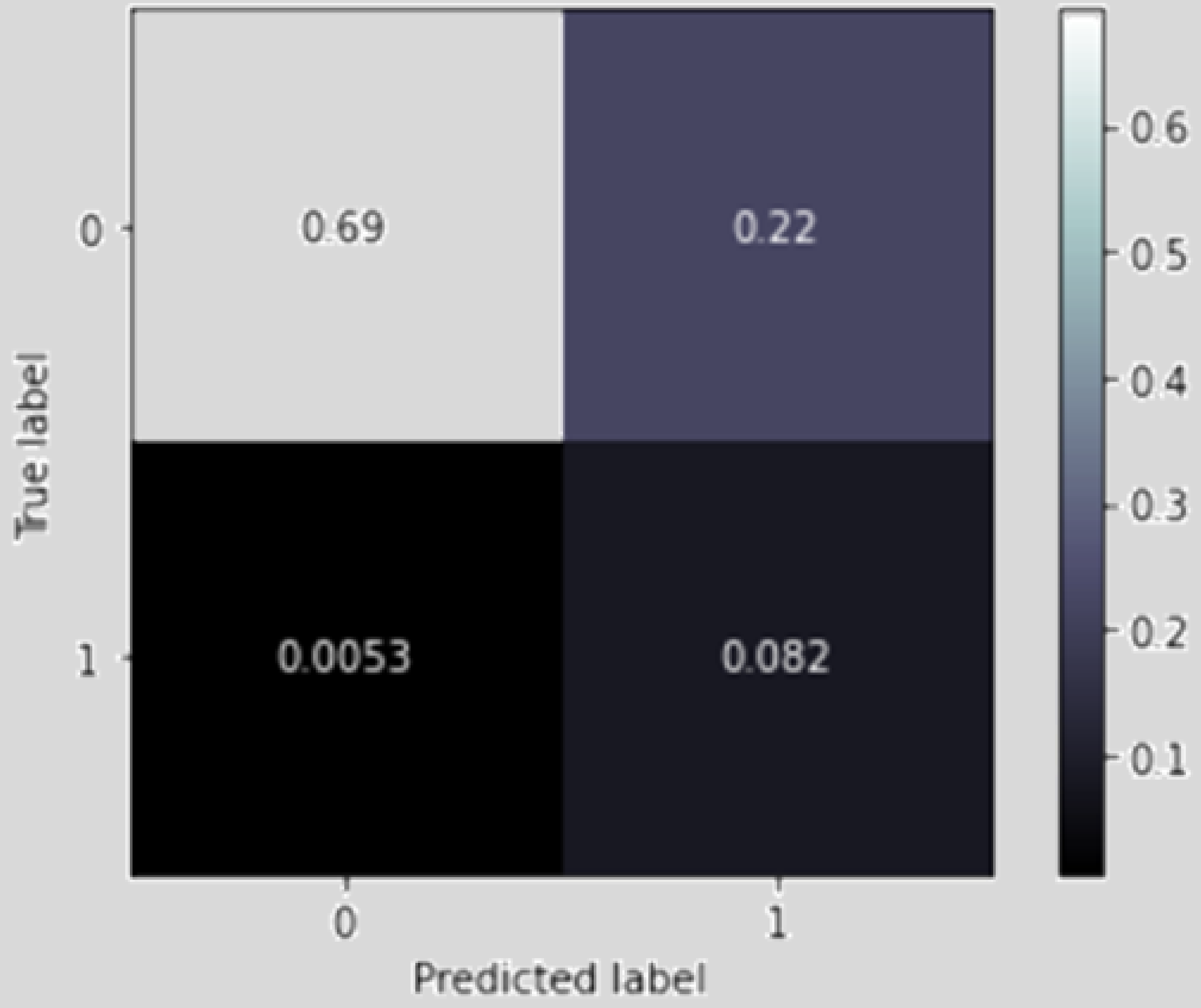
	Methodology	Rationable
Data cleaning	Remove records	Missing values constitutes < 1% of total column
Clustering	• Multi-feature clustering • K-Means algorithm • Boruta feature selection & highly correlated variables removal	Reduce dimensionality of the dataset
Predictive modelling	• Train: test split = 70: 30 • XGB classifier • RandomGridSearchCV • Boruta feature selection & highly correlated variables removal • SMOTE resampling • Target: "TURNDOWN"	• XGB is a widely known model for its ability to return accurate results • SMOTE is used to overcome data imbalance

Environment: Jupyter Lab
Programming language: Python
Dataset: Survey of Consumer Finances 2010, 2013, 2019 (US)

Model results

Using XGBClassifier the model returns a test accuracy of 77%

Confusion matrix for test data set



<u>Train data set</u>

Accuracy: 87%

Specificity: 0.77

Sensitivity: 0.98

<u>Test data</u>

Accuracy: 77%

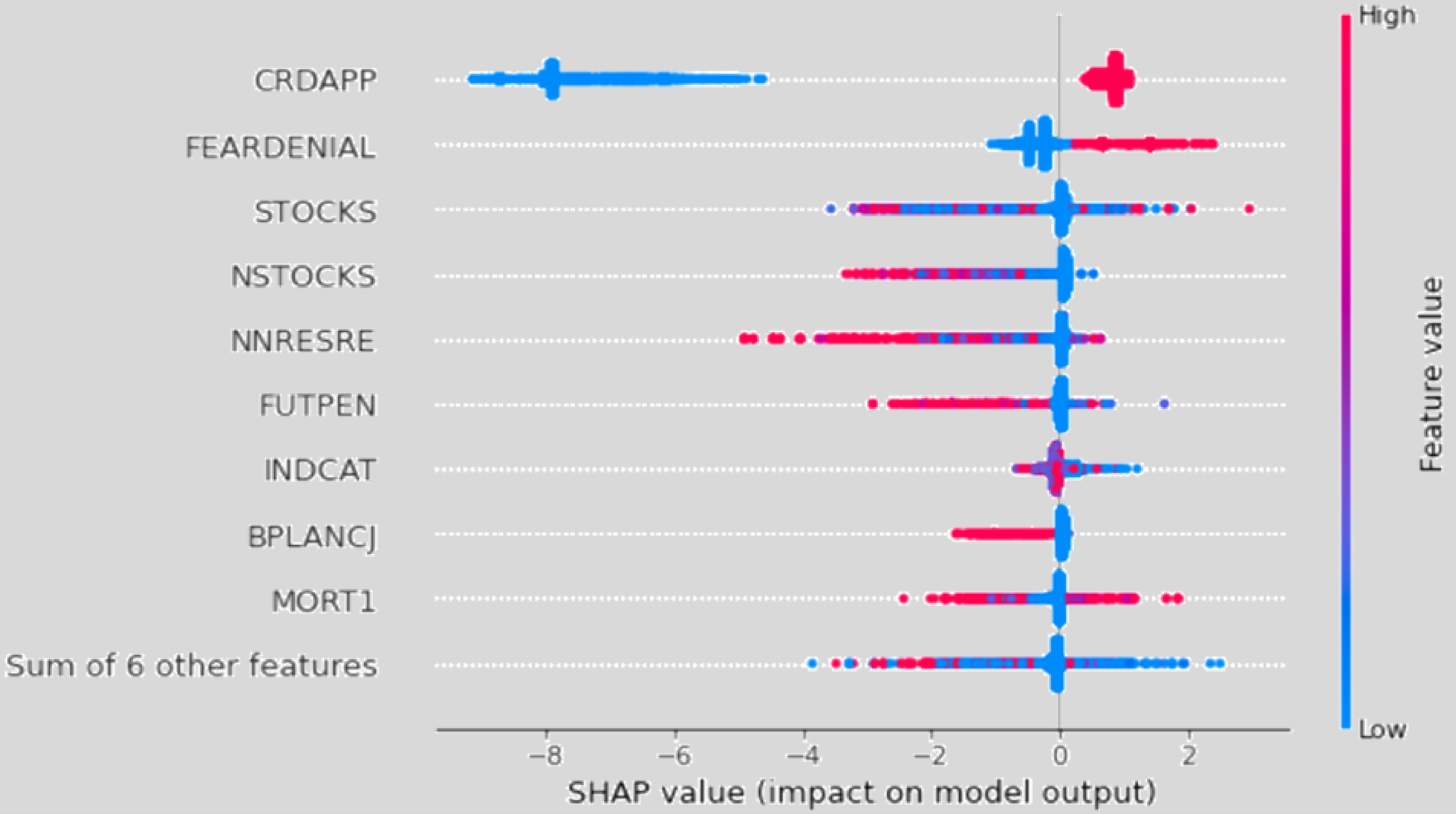
Specificity: 0.77

Sensitivity: 0.94

Model best parameters: {'reg_lambda': 0.5, 'reg_alpha': 1, 'n_estimators': 100, 'min_child_weight': 2, 'max_depth': 8, 'gamma': 7}

Feature importance

15 features were selected and fed into the prediction model

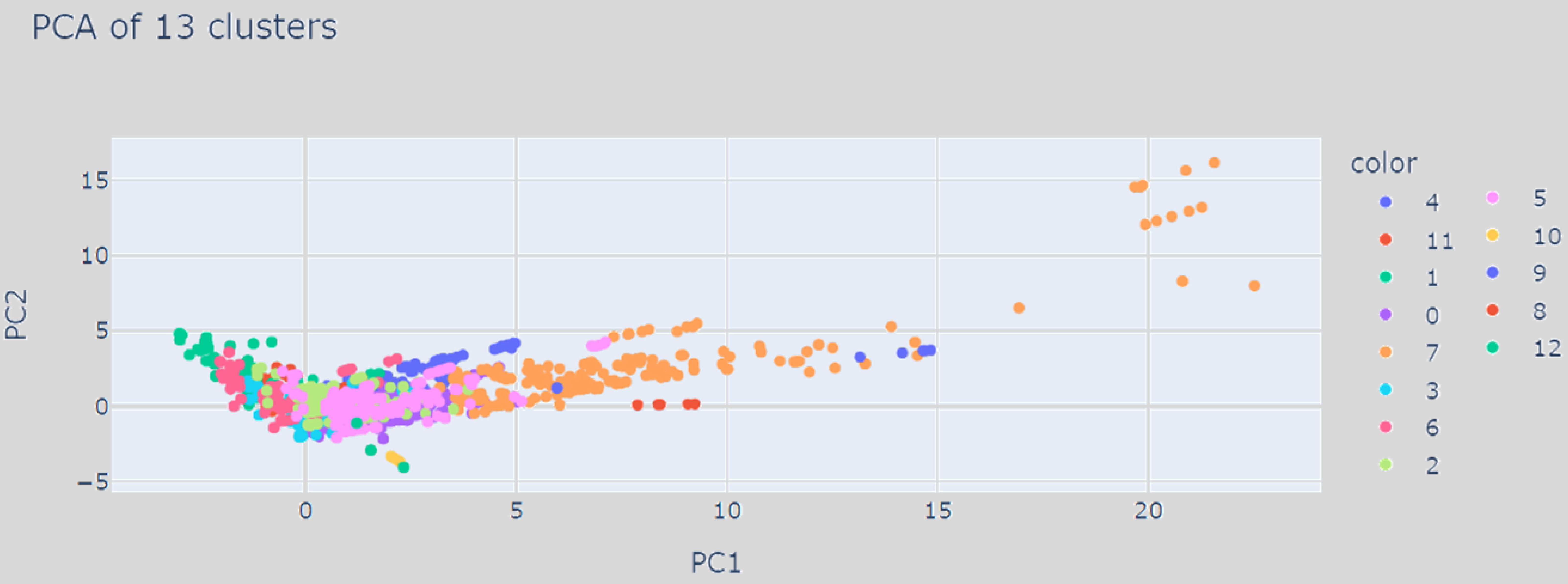


Features selected from Boruta algorithm and removal of high correlated features:

'INDCAT', 'LATE60', 'CRDAPP', 'FEARDENIAL', 'EMERGCUT', 'CKRECOMFRND', 'CKPERSONAL', 'STOCKS', 'NSTOCKS', 'FUTPEN', 'BPLANCJ', 'SAVBND', 'NNRESRE', 'MORT1', 'LLOAN4'

Cluster analysis

13 customer groups were generated through using K-means clustering



Visualization of 13 clusters generated K-Means and summarized using PCA

Descriptive analysis

Younger age groups are more likely to be turned down from credit approvals