Table of Contents

1. Executive Summary

In the first section, we performed a cluster analysis (CA) using a sample of 500 borrowers in Lending club data to identify the factors affecting each variable, and then cluster those factors into informative characteristics of each cluster, which provides management with insights on customer behaviour. Based on the data, we were able to identify the 3 key clusters. Cluster 1 has borrowers with high loans, payments, interest rates, and high sub grade, low debt-to-income ratios. Cluster 2 has higher interest rates and high sub grades indicating moderate risk. Cluster 3 has high dti rates, but negative values for other factors.

In the second section, a controlled experiment on a loan discrimination case was designed to examine how interactions between people and models might impact human judgments. We created four conditions, and each condition was conducted with 10 participants. We performed the Kruskal-Wallis and Mann-Whitney U-tests to see whether different conditions would affect participants' decisions. We also used ANOVA to see whether different conditions affect each participant's confidence. After testing and comparing the results, we concluded that model prediction leads to worse accuracy and does not affect participants' confidence.

2. Cluster Analysis

2.1 Introduction

Cluster analysis was used to group borrowers with similar features for marketing and product management decisions in Lending Club, a peer-to-peer lending company. The dataset had 50,000 observations and 53 variables, with some null values dropped before analysis.

2.2 Data preparation

We examined all variables and did not use categorical variables for CA and excluded them. We used total payment, loan amount, total received principal, instalment amount, total received interest, debt to income rate, sub grade, last payment amount, interest rate, employment length, annual income, open accounts, revolt balance, total credit received and total collection amount variables to conduct CA (Table A.1 in Appendix).

Mahalanobis distance was used for outlier detection, and observations with higher distances were removed if their p-values were less than 0.001. Standardization was used to eliminate scale differences between clustering variables. Multicollinearity among variables was checked, as it can over-weigh constructs, and equal weight among variables is desired. Results showed strong relationships between some variables, which can affect cluster solutions. (Figure A.1). Thus, we conducted Principal Component Analysis and Factor Analysis. To check the suitability of the data set for PCA and FA we conducted the Kaiser-Meyer-Olkin test and The Bartlett test. KMO is 0.67 which is greater than 0.50. The Bartlett test result is $p<0.05$. PCA results showed cross loading and missing variable, hence we used FA to rotate the components. In our case, PC extraction with Orthogonal rotation with a 4 factors solution is the best result. Therefore, we dropped 7 variables.

2.3 Clustering Analysis and results

We used Hierarchical clustering algorithm and K-means to perform cluster analysis. We ran a function which gave us the best linkage method as Ward, thus this method was used for clustering.

We used gap statistics (Figure 1.) and dendrogram (Figure 2.) to determine the number of clusters, the gap statistic result was hard to interpret and thus we relied on the dendrogram. Based on the dendrogram, 3 clusters were chosen as they have a decent distance between them.

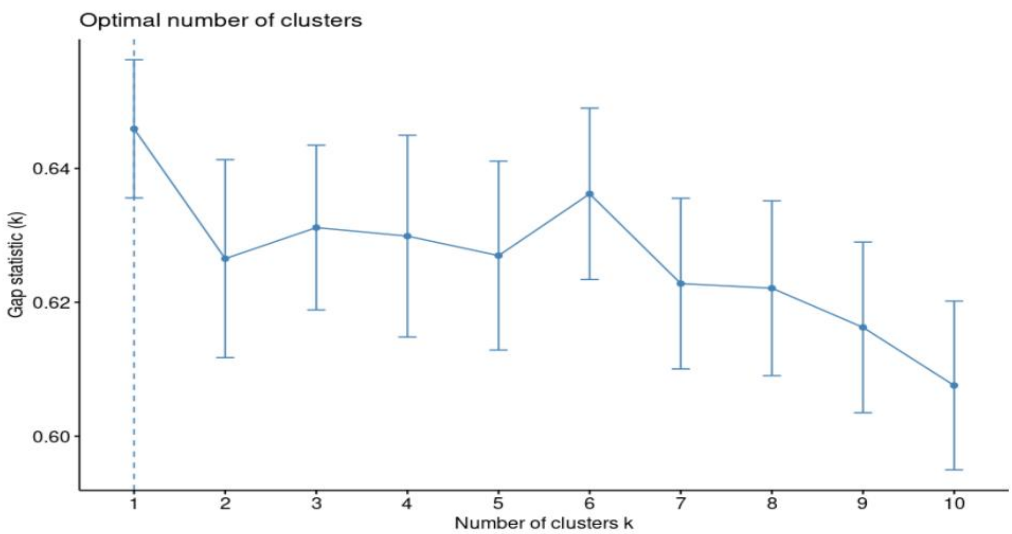

Figure 1. Optimal number of clusters

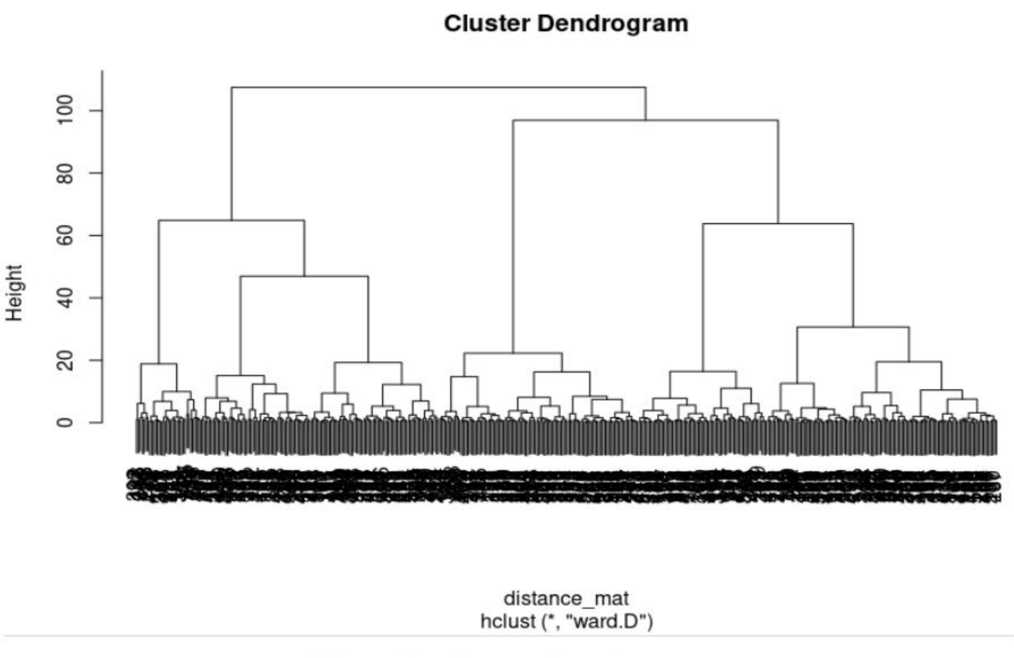

Figure 2. Cluster Dendrogram

The measure of similarity between the observations being clustered is a crucial decision. We used Euclidean distance to find the distance matrix. Figure 3. below shows the mean values for each cluster.

cluster	RC1	RC2	RC3	RC4	cluster
1	0.9358614	0.1262009	0.45810588	-0.04481272	1
2	-0.5502258	0.1618964	-0.34480517	-0.55821338	2
3	-0.4331947	-0.4842687	-0.07592393	1.07513696	3

Figure 3. Mean values of clusters

The first cluster has negative values in RC4 factor which shows debt to income rate, and high positive value for other factors which includes total payment, loan amount, total received principal, instalment amount, and total received interest (RC1), interest rate and sub grade (RC2), total current amount and annual income (RC4). This shows that cluster 1 contains borrowers with high loan amounts and total payments, both in principal and interest.

The second cluster has positive values in RC2 factor, which includes interest rate and sub grade meaning the borrowers belonging to this cluster have higher interest rates and higher sub grade. This cluster has low values in all other factors.

The third cluster has high positive values in RC4 factor, which includes debt to income rate and negative values for other factors. This suggests that this cluster has borrowers with high dti rate.

Overall, these clusters can help to identify distinct groups of borrowers with distinct characteristics and behaviour. Cluster 1 has borrowers who have taken higher loans and made higher payments with higher interest rate and high sub grade however these borrowers' debt to income ratio is low. Loans in this cluster may be characterized by borrowers with higher creditworthiness or better financial stability. Cluster 2 has borrowers with higher interest rates and high sub grades and lower in dti and other factors which makes this cluster moderate group. Cluster 3 has borrowers with low current balance and annual income, but higher debt-to-income ratios. Loans in this cluster may have higher risk or lower creditworthiness than other loans.

To validate the clusters, we ran the silhouette score to identify the cluster boundaries. That would help us in ensuring that the clusters are significant. For the K-means clustering we got a mean silhouette score of 0.22 and for the hierarchical clustering we got a mean silhouette score of 0.23 which both indicate that the clustering was good, having a good fit with the clusters assigned (Figure A.2.).

These clusters can be used by the company to target specific types of customers based on clusters 1, 2, 3. The product managers can create or change specific products for the different sets of “clusters” which can drive the business of the company.

3. Simple Controlled Experiment

3.1. Introduction

We conducted a controlled experiment on loan discrimination to examine how interactions between people and AI (Artificial Intelligence) models can affect human judgments. We created several hypotheses and followed the principles of Randomized Field Trials and Blind Tests in the experimental design to ensure a rigorous investigation. We also considered participants' backgrounds in the survey questions to counteract accuracy float and structured the questions differently for each condition.

3.2. Research Hypotheses

We proposed two research hypotheses based on the assumptions that AI predictions perform better than manual predictions and decision-makers will adjust their own predictions based on AI predictions:

Hypothesis 1 People who used the algorithm should be able to make more accurate predictions than they would otherwise

Hypothesis 2 People should have a higher level of confidence in performing a more accurate prediction when machine learning predictions are presented

3.3. Experimental Setup

We randomly selected 10 loan applicants from a dataset of 40 to evaluate risk assessments. Participants were presented with a numerical description and machine learning predictions through Qualtrics to test their interaction with the assessments. All applicants were assumed to apply for the same loan type and interest rate.

Our experiment survey included a consent page, a tutorial, an intro survey, the primary experimental task where participants made predictions on loan applicants, and an exit survey. (Appendix 6.1).

Intro survey: To counteract the accuracy float, we asked participants about their backgrounds at the start of each survey. We hypothesized that age, education, and job relevance could impact accuracy. Participants with more experience and knowledge may make more accurate decisions by considering relevant factors.

Primary experimental task: Participants had to approve or reject a loan based on narrative profiles and rate their confidence level on a 0-100% scale. The profiles included loan ID, amount, annual income, credit score, credit history, and risk assessment. Figure 4. shows different question structures based on different conditions: Baseline, RA Prediction, Update, and Feedback. Participants were randomly assigned to one of four conditions but were blinded to the factors being tested and their specific conditions.

<table>
<tr><td colspan="2" rowspan="4">Information Provided</td><td colspan="4">Feedback Provided</td></tr>
<tr><td colspan="2">No</td><td colspan="2">Yes</td></tr>
<tr><td colspan="2">Decision Update</td><td colspan="2">Decision Update</td></tr>
<tr><td>No</td><td>Yes</td><td>No</td><td>Yes</td></tr>
<tr><td rowspan="2">Predictions Provided</td><td>No</td><td>BASELINE</td><td></td><td></td><td></td></tr>
<tr><td>Yes</td><td>RA PREDICTION</td><td>UPDATE</td><td>FEEDBACK</td><td></td></tr>
</table>

Figure 4. Four conditions in the experiment

Exit survey: In the exit survey (excluding baseline participants), we asked participants about their thoughts on the computer algorithm's accuracy and their reliance on it. We also asked about the most significant variables and if they needed more information to make better decisions in future studies.

3.4. Analysis & Results

We collected 53 valid responses through random questionnaires sent to family, friends, and schoolmates. Table 1. displays descriptive statistics of the collected data attributes.

	Baseline N=17	RA Prediction N=13	Update N=12	Feedback N=11	Total N=53
Demographics					
18-25 years old	82.40%	92.30%	91.70%	63.60%	83.02%
26-40 years old	17.60%	7.70%	8.30%	27.30%	15.09%
41-60 years old	0.00%	0.00%	0.00%	9.10%	1.89%
High school diploma/GED	0.00%	15.40%	0.00%	0.00%	3.77%
Bachelor's degree	47.10%	23.10%	58.30%	27.30%	39.62%
Master's degree or higher education	52.90%	61.50%	41.70%	72.70%	56.60%
Finance or data process industry	23.50%	23.10%	0.00%	18.20%	16.98%
Outcomes - final decisions					
Type I error rate (AI: 0.28)	0.20	0.20	0.18	0.16	0.19
Type II error rate (AI: 0.33)	0.20	0.41	0.36	0.15	0.28
Average Confidence Level	69.28%	66.55%	73.00%	70.03%	69.61%

Table 1. Attributes of the participants in our experiments

We analyzed participants' performances in the final stage using three outcome metrics: prediction accuracy (Type I and II error rates) and confidence in their predictions (Average Confidence Level).

3.4.1. Type I error rate

We examined the Type I error rate (rejecting a good loan) as the first outcome variable. The histograms indicated non-normal distribution across all four conditions, so we used the Kruskal Wallis test for comparisons and the Mann-Whitney U test for pairwise comparisons. The results show that all p-values are not significant. In other words, there is no significant difference in Type I under different conditions, which can also be seen intuitively in the barplot (Figure 5.) and the boxplot (Figure 6.). One explanation is that when participants are not sure of the correctness of their judgments, they tend to choose approval, and thus there are fewer cases of wrongly rejecting good loans, so it is difficult to reflect the corrective effect of AI.

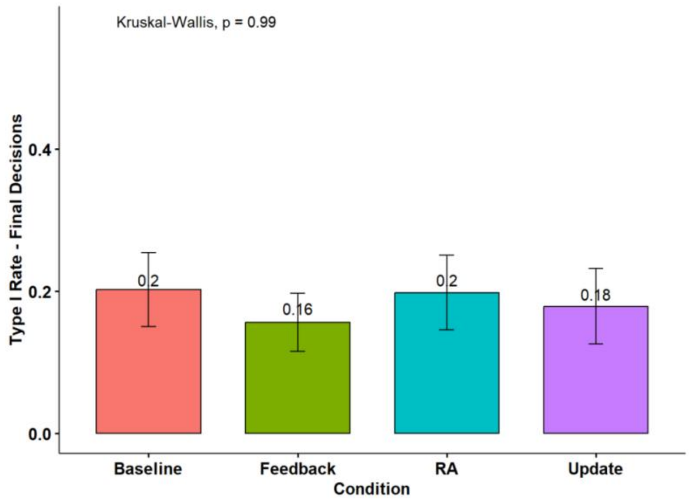

Figure 5. Barplot of Type I error rate for each condition

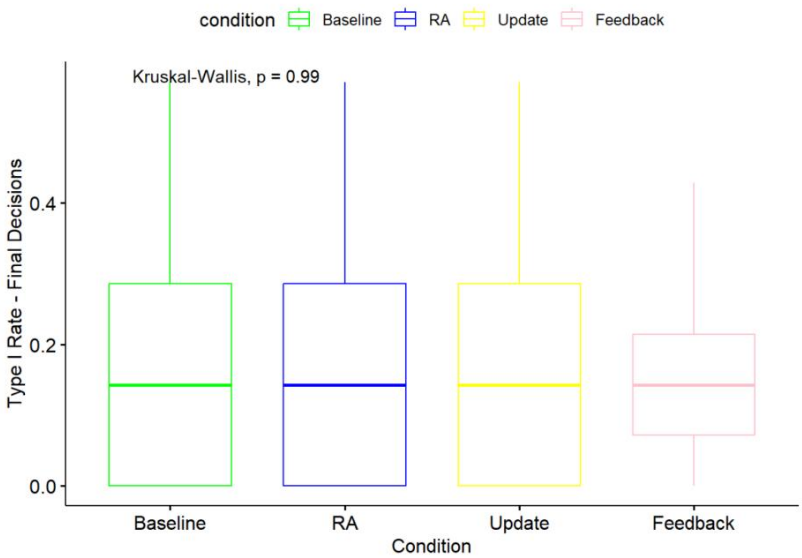

Figure 6. Boxplot of Type I error rate for each condition

3.4.2. Type II error rate

Then we examined the Type II error rate (Type II), which represents the probability of approving a bad loan. The Kruskal Wallis test showed a significant difference in Type II across conditions. Further analysis using the Mann-Whitney U test showed that the mean of Type II for RA (Research Assistant) was larger than the means of Type II for Baseline at a significance level of 0.1, indicating a negative impact of AI on Type II. However, Feedback performed better than RA at a significance level of 0.1 and was not significantly different from Baseline, suggesting that providing actual outcomes of loans could be an effective training method that compensates for the distraction caused by AI. The visualization (Figure 7.&8.) of the result is as follows.

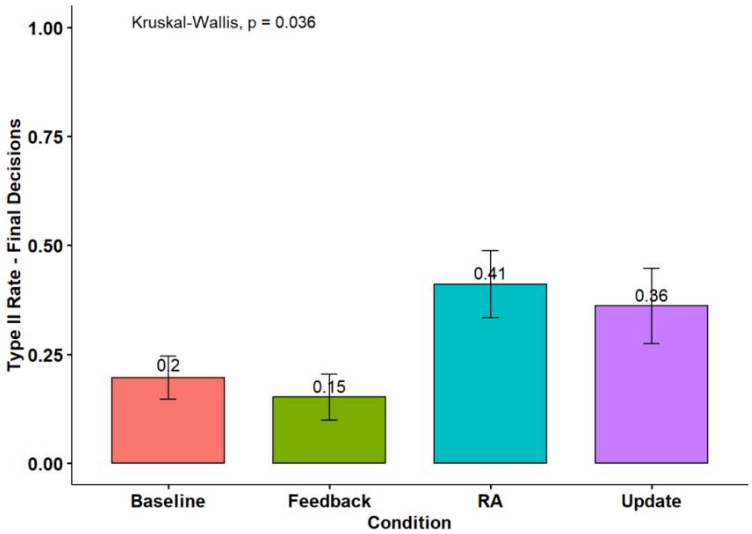

Figure 7. Barplot of Type II error rate for each condition

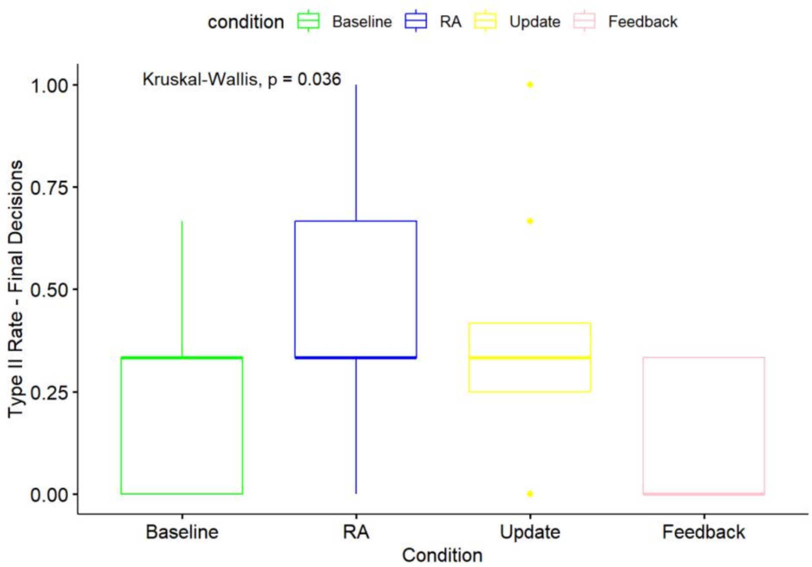

Figure 8. Boxplot of Type II error rate for each condition

3.4.3. Average Confidence Level

We looked at the Average Confidence Level (Confidence) variable, which is the average of each participant's confidence in their predictions. We confirmed that the data met the requirements for ANOVA, and then used it to compare the four conditions while controlling for age, education, and related work experience. The results showed that, except for education, there were no significant differences between the conditions. The visualization (Figure 9.&10.) of the result is as follows:

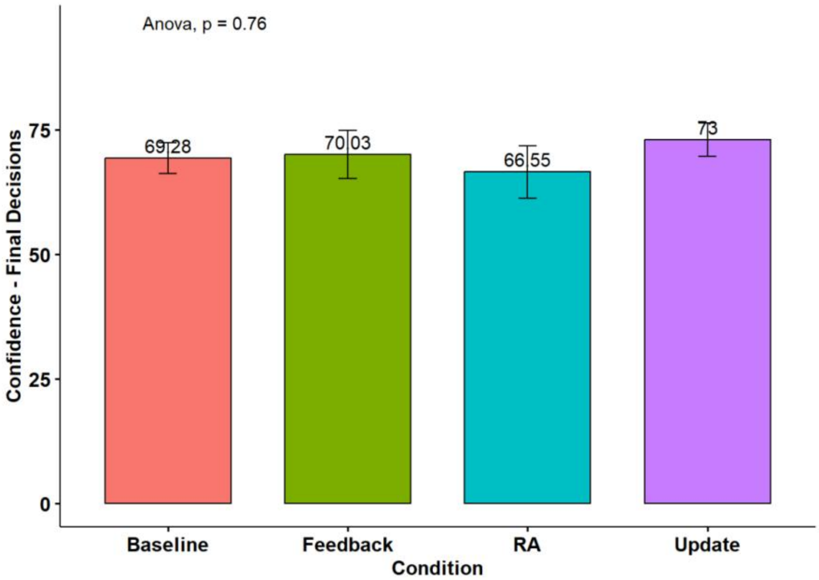

Figure 9. Barplot of average confidence for each condition

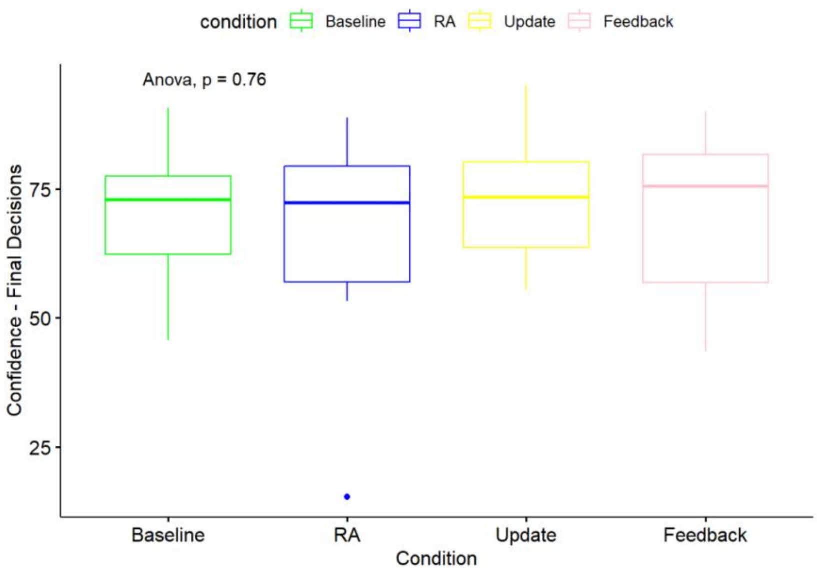

Figure 10. Boxplot of average confidence for each condition

3.5. Discussion

Based on the above results, we find that both of our hypotheses should be rejected. Moreover, we could draw the following conclusions:

(1) AI predictions appear to lead to worse accuracy: Participants' rejection of good loans is independent of whether they were presented with AI predictions, while they are more likely to mistakenly approve bad loans when assisted by AI.

(2) AI predictions do not affect participants' confidence in their predictions.

To a certain extent, we recommend that managers avoid using AI, but decision-makers can be trained by repeatedly presenting real outcomes. Nevertheless, more experiments are needed to verify this conclusion, considering the limitations of our experiment: a small sample size, low diversity of participants' backgrounds, lack of incentives to make the right choices, and low AI accuracy (70%).

4. Conclusion

In conclusion, Cluster analysis identified three distinct groups of borrowers with distinctive characteristics and behavior. The clusters were validated using silhouette score and can be used to target specific types of customers and drive the business of the company.

Another discovery we found was that AI predictions tend to result in poorer judgments in terms of accuracy and have no significant effect on people's confidence. The limitations of our experiment are the small sample size and limited variety.

5. Appendix

5.1. Tables and Figures

Variable name	Description	Data Type	Status
id	A unique LC assigned ID for the loan listing	Numeric	Not used
member_id	Member idfentifier	Numeric	Not used
loan_amnt	Loan amount	Numeric	Used
funded_amnt	Funded amount is the same with loan amount	Numeric	Not used
funded_amnt_inv	The same with loan amount	Numeric	Not used
term	46K observation with36 term	Numeric	Not used
int_rate	Interest rate	Numeric	Used
installment	installment	Numeric	Used
grade	Categorical ordinal 7 classes	Categorical	Not used
sub_grade	Categorical ordinal 35 classes	Categorical	Used
emp_title	Text data	String	Not used
emp_length	Length of employment	Numeric	Used
home_ownership	Categorical	Categorical	Not used
annual_inc	Annual income	Numeric	Used
verification_status	Categorical	Categorical	Not used
issue_d	Date	Date	Not used
loan_status	Categorical	Categorical	Not used
pymnt_plan	Categorical Nominal	Categorical	Not used
desc	Text data	String	Not used
purpose	Categorical	Categorical	Not used
title	Text data	String	Not used
zip_code	Address don't give info about cluster	Text	Not used
addr_state	Address don't give info about cluster	Text	Not used
dti	Debt to income rate	Numeric	Used
delinq_2yrs	Deliquency coltains 42k observations 0	Numeric	Not used
earliest_cr_line	Date	Date	Not used
inq_last_6mths	25K observations with zero	Numeric	Not used
mths_since_last_delinq	Numeric with a lot of missing value (28K)	Numeric	Not used

Variable name	Description	Data Type	Status
mths_since_last_record	Numeric with a lot of missing value (47K)	Numeric	Not used
open_acc	Oppen account	Numeric	Used
pub_rec	Non distinct value 0	Numeric	Not used
revol_bal	Total credit revolving balance	Numeric	Used
revol_util	Revolv utility	Numeric	Used
total_acc	Total accounts	Numeric	Used
total_pymnt	Total payment	Numeric	Used
total_pymnt_inv	Repeated values with TP	Numeric	Not used
total_rec_prncp	Numeric	Numeric	Used
total_rec_int	Numeric	Numeric	Used
total_rec_late_fee	48K observations are 0	Numeric	Not used
recoveries	44K observations are 0	Numeric	Not used
collection_recovery_fee	45K observations are 0	Numeric	Not used
last_pymnt_d	Date	Date	Not used
last_pymnt_amnt	Numeric	Numeric	Used
next_pymnt_d	Date	Date	Not used
last_credit_pull_d	Date	Date	Not used
collections_12_mths_ex_med	49K observations are 0	Numeric	Not used
mths_since_last_major_derog	42K observations are 0	Numeric	Not used
policy_code	Non distinct value 1	Numeric	Not used
acc_now_delinq	49K observations are 0	Numeric	Not used
tot_coll_amt	Numeric	Numeric	Used
tot_cur_bal	Numeric	Numeric	Used
total_credit_rv	Numeric	Numeric	Used
loan_is_bad	Categorical Nominal	Categorical	Not used

Table A.1 Choice of variables

	sb_gr	dti	rvl_t	ttl_c	ttl_p	lst__	tt_cr_	ln_mn	int_r	instl	emp_l	annl_	opn_c	rvl_b	ttl_rc_p	ttl_rc_n	ttl_c_	tt_cl_
sub_grade	1.00																	
dti	-0.14	1.00																
revol_util	-0.38	0.24	1.00															
total_acc	-0.03	0.23	-0.06	1.00														
total_pymnt	-0.25	0.02	0.10	0.23	1.00													
last_pymnt_amnt	-0.13	-0.04	-0.01	0.15	0.49	1.00												
tot_cur_bal	0.07	-0.02	0.04	0.34	0.27	0.15	1.00											
loan_amnt	-0.29	0.04	0.10	0.26	0.89	0.39	0.29	1.00										
int_rate	-0.98	0.15	0.41	0.02	0.22	0.12	-0.09	0.26	1.00									
installment	-0.27	0.03	0.13	0.24	0.89	0.36	0.26	0.96	0.25	1.00								
emp_length	-0.02	0.04	0.05	0.13	0.11	0.04	0.10	0.12	0.02	0.11	1.00							
annual_inc	0.01	-0.16	0.02	0.17	0.25	0.13	0.42	0.26	-0.03	0.25	0.06	1.00						
open_acc	-0.05	0.30	-0.10	0.66	0.16	0.07	0.24	0.19	0.04	0.18	0.04	0.11	1.00					
revol_bal	-0.02	0.14	0.17	0.22	0.28	0.12	0.46	0.31	0.01	0.29	0.09	0.31	0.22	1.00				
total_rec_prncp	-0.05	-0.02	0.04	0.21	0.95	0.58	0.26	0.78	0.04	0.82	0.09	0.24	0.15	0.26	1.00			
total_rec_int	-0.58	0.11	0.21	0.16	0.71	0.08	0.15	0.74	0.55	0.65	0.10	0.14	0.13	0.19	0.46	1.00		
total_credit_rv	0.14	0.06	-0.18	0.30	0.29	0.15	0.43	0.32	-0.17	0.29	0.09	0.29	0.33	0.85	0.30	0.13	1.00	
tot_coll_amt	-0.02	-0.02	-0.03	0.02	-0.02	-0.01	0.00	-0.02	0.02	-0.02	0.00	0.00	0.01	-0.02	-0.02	-0.01	-0.02	1.00

Figure A.1. Correlation matrix

```
> K_sil <- silhouette(km$cluster, dist(fscores))
> summary(K_sil)
Silhouette of 465 units in 3 clusters from silhouette.default(x = km$cluster, dist = dist(fscores)) :
 Cluster sizes and average silhouette widths:
      267       122        76
0.2806666 0.1451216 0.1592006
Individual silhouette widths:
    Min.  1st Qu.  Median     Mean  3rd Qu.     Max.
-0.08168  0.13914  0.23104  0.22525  0.32797  0.47657
>
>
> H_sil <- silhouette(final_data$cluster, distance_mat)
> summary(H_sil)
Silhouette of 465 units in 3 clusters from silhouette.default(x = final_data$cluster, dist = distance_mat) :
 Cluster sizes and average silhouette widths:
       84       307        74
0.1370243 0.2762564 0.1488967
Individual silhouette widths:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.2480  0.1383  0.2502  0.2308  0.3414  0.4757
Session restored from your saved work on 2023-Mar-17 15:23:04 UTC (1 day ago)
>
```

Figure A.2. Silhouette score (validation)

5.2. Link to the Experiment and Screenshots of the Experiment Interface

Experiment link: https://wbs.qualtrics.com/jfe/form/SV_b7SZjcai1JmTcIm

Screenshots:

(1) Intro survey

4. What is your work industry?

- ◯ Agriculture
- ◯ Finance
- ◯ Entertainment
- ◯ Education
- ◯ Health care
- ◯ Data processing
- ◯ Legal services
- ◯ Military
- ◯ Others
- ◯ Prefer not to say
- ◯ Currently not working

(2) Baseline

Applicant profile
Loan applicant #3157804 has applied a loan amount of $25000. The applicant has inquired 2 other credit inquiries in the last 6 months. The applicant has an annual income of $82000 and a A5 loan grade (A is the best loan grade (least risky), 1 is the best subgrade).

Make a prediction:
Please make your decision for the loan.

- ◯ Approve
- ◯ Reject

How confident are you in your decision? (0: very unconfident; 100: very confident)

0 10 20 30 40 50 60 70 80 90 100

Confidence

(3) RA Prediction

Applicant profile

Loan applicant #3157804 has applied a loan amount of $25000. The applicant has inquired 2 other credit inquiries in the last 6 months. The applicant has an annual income of $82000 and a A5 loan grade (A is the best loan grade (least risky), 1 is the best subgrade).

Risk assessment:

According to the risk score algorithm, the applicant should be approved.

Make a prediction:

Please make your decision for the loan.

○ Approve

○ Reject

How confident are you in your decision? (0: very unconfident; 100: very confident)

0 10 20 30 40 50 60 70 80 90 100

Confidence

(4) Update

Applicant profile

Loan applicant #3157804 has applied a loan amount of $25000. The applicant has inquired 2 other credit inquiries in the last 6 months. The applicant has an annual income of $82000 and a A5 loan grade (A is the best loan grade (least risky), 1 is the best subgrade).

Make a prediction:

Please make your decision for the loan.

○ Approve

○ Reject

How confident are you in your decision? (0: very unconfident; 100: very confident)

0 10 20 30 40 50 60 70 80 90 100

Confidence

（Next page)

Risk assessment:

According to the risk score algorithm, the applicant should be approved.

Make a prediction again:

○ Approve

○ Reject

How confident are you in your decision? (0: very unconfident; 100: very confident)

0 10 20 30 40 50 60 70 80 90 100

Confidence

(5) Feedback

Applicant profile

Loan applicant #3157804 has applied a loan amount of $25000. The applicant has inquired 2 other credit inquiries in the last 6 months. The applicant has an annual income of $82000 and a A5 loan grade (A is the best loan grade (least risky), 1 is the best subgrade).

Risk assessment:

According to the risk score algorithm, the applicant should be approved.

Make a prediction:

Please make your decision for the loan.

- ○ Approve
- ○ Reject

How confident are you in your decision? (0: very unconfident; 100: very confident)

0 10 20 30 40 50 60 70 80 90 100

Confidence

(Next page)

The actual outcome of the loan is good.