Monitoring and Governance

By Urielle fotso

MODEL MONITORING

- Initial Model Fit Statistics

- 1.1. Model Drift Monitoring

- PSI or Chi-Square Example Using Build and Holdout

- Risk Tiering

- 2.0. Variable Level Monitoring

- 2.1. Model Build Variable Level Statistics for Every Input

- 2.2. Acceptable Ranges

- 2.3. Missing Values

- 2.4. Variable Drift Monitoring

- CSI or Chi-Square

- Risk Tiering

1.0. Initial Model Fit Statistics
• 1.1. Model Drift Monitoring

Target 1 (fully recovered) and target 2 (partially recovered)

	expected	actual	O-E	(O-E)^2	(O-E)^2/E			
Target 1	296							
Target 2	154							
total	450							
					3.09	Chi square		
					3.98	Chi square(p=0.05) for df=1		
					6.54	Chi square (p=0.025) for df=1		
					7.98	Chi square (p=0.01) for df=1		

Having this chi square at 4.13 we are at least 95% sure that the model has shifted but not 97% or 99%.

• Risk Tiering

Based on the Chi-square values at different confidence intervals:

- Chi-square = 4.13 (95% confidence interval): The appropriate risk tiering is to take no immediate action but monitor closely, as the drift is slight and manageable.

- Chi-square = 6.54 (97% confidence interval): This indicates a significant drift. Given the sensitivity of loan-related projects, the model should be stopped and rebuilt.

- Chi-square = 7.98 (99% confidence interval): This reflects a substantial drift, and the appropriate action is to rebuild the model immediately.

• 2.0. Variable Level Monitoring
• 2.1. Model Build Variable Level Statistics for Every Input

My best model for the prediction of fully and partially recovered was forward regression and the selected features were ag , number of dependents, interest rate , days past due, collection attempts ,loan type, payment history and collection methods . Below is the distribution statistic of interval variables;
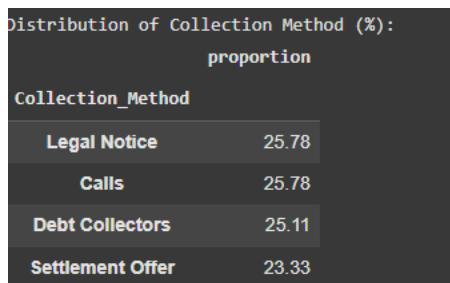
	Age	Num_Dependents	Interest_Rate	Days_Past_Due	Collection_Attempts
count	450.000000	450.000000	450.000000	450.000000	450.000000
mean	42.895556	1.500000	11.193000	70.071111	2.713333
std	12.763402	1.139003	3.789908	60.127958	2.596906
min	21.000000	0.000000	5.020000	0.000000	0.000000
25%	32.000000	0.000000	7.865000	3.250000	1.000000
50%	44.000000	1.000000	11.000000	64.000000	2.000000
75%	53.000000	3.000000	14.540000	121.750000	3.000000
max	64.000000	3.000000	17.960000	180.000000	10.000000

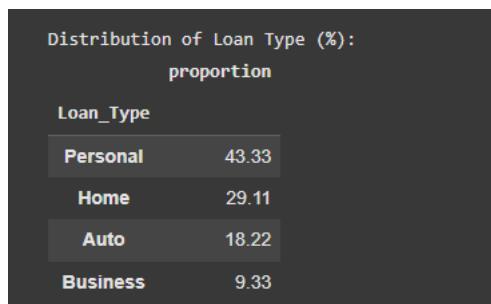The distribution for the categorical variable are as follows ;

Proportion of Payment history

Distribution of Payment History (%):
                proportion

Payment_History

On-Time	61.33
Delayed	27.78
Missed	10.89

Proportion of collection method variables

Distribution of Collection Method (%):
                proportion

Collection_Method

Legal Notice	25.78
Calls	25.78
Debt Collectors	25.11
Settlement Offer	23.33

Proportion of loan type

Distribution of Loan Type (%):
                proportion

Loan_Type

Personal	43.33
Home	29.11
Auto	18.22
Business	9.33

2.2. Acceptable Ranges (Caps & Floors)

- Age: 18–64; outside this range → cap and floor or sub model.

- Number of Dependents: 0–9; outside this range → cap and floor.

- Interest Rate: 2%–24%; outside this range → cap and floor.

- Days Past Due: No cap and floor; all values accepted.

- Collection Attempts: >20 → cap/floor

2• 3. Missing Values (Imputation Values)

1. Age: Impute with median; >64 or <21 → segment population and reassess feature importance.

2. Number of Dependents: Impute with median; above max → reassess model; below min → set to 0.

3. Interest Rate: Impute with median; out-of-range → cap/floor.

4. Days Past Due: Impute with 0 or 64; <0 → set to 0; >max → set to max.

5. Collection Attempts: Impute with median; <0 → set to 0; >10 → cap to 10.

6. Collection Method: Impute with mode.

7. Loan Type: Impute with mode (most frequent).

8. Payment History: Impute with mode (most frequent).

• 2.4. Variable Drift Monitoring

• CSI for numerical variable

i)   age

age bin	age	# expected	#actual	% expected	%actual	index|PSI
1	21					0.16
2	31					0.0098
3	41					0.0055
4	51					0.0002
5	61					0
6	71					0.0018
total						0.177

The CSI is between 0.1 and 0.2 implying some minor changes. For risk tiering, nothing will be done

ii)   Number of dependent

Number of dependent	# expected	#actual	% expected	%actual	Index |PSI
1					0.16
2					0.0098
3					0.0055
4					0.0002
5					0.04
6					0.0018
Total					0.2173

The CSI is greater than 0.2 meaning significant change, for risk tiering we will refit.

iii)   interest rate

Interest rate	# expected	#actual	% expected	%actual	Index | PSI
5					0.16
6					0.02
7					0.0098
8					0.0001
9					0.006
10					0.005
11					0.0055
12					0.0003
13					0.0001
14					0.0002
15					0.0005
16					0
17					0.04
Total					0.2475

The CSI is greater than 0.2 meaning significant change, for risk tiering we will rebuild.

iv)   days past due

Days past due	# expected	#actual	% expected	%actual	Index | PSI
5					0.16
7					0.0098
11					0.0055

14					0.0002
17					0.04
Total					0.2155

The CSI is greater than 0.2 meaning significant change , for risk tiering we will rebuild

v) collection attempts

Collection attempts	# expected	#actual	% expected	%actual	index|PSI
0					0.16
1					0.0098
2					0.0055
3					0.0002
4					0.04
5					0.0004
6					0.0003
7					0.00001
8					0.0004
9					0.00003
10					0.00001
total					0.2166

The CSI is greater than 0.2 meaning significant change , for risk tiering we will rebuild

Chi-Square for categorical features

i) loan type

Loan type	Expected	actual	O-E	(O-E)^2	(O-E)^2/E			
Education								
Auto								
Home								
Business								
Total	450							
					5.09	Chi square		
					6.13	chi square(p=0.05) for df=3		
					6.54	Chi square (p=0.025) for df=3		
					7.98	Chi square (p=0.01) for df=3		

At 95% confidence interval there is a drift to 6.13 but this not strong enough to cause a significant change taken into account the degree of freedom. so for risk tiering nothing will be done . The chi square at 97.5% and 99% are quiete drifting , so for the risk tiering we will refit the model.

     ii)   payment history

	Expected	actual	O-E	(O-E)^2	(O-E)^2/E			
Payment history	450							
Missed								
Delayed								
Ontime								
Total	450							
					1.009	Chi square		
					1.13	Chi square(p=0.05) for df=3		
					2.54	Chi square (p=0.025) for df=3		
					2.98	Chi square (p=0.01) for df=3		

At 95% confidence interval there is a drift to 1.13 but this not strong enough to cause a significant change. so for risk tiering nothing will be done. The chi square at 97.5% and 99% are alarming , so for the risk tiering we will rebuild the model

     iii)   collection method

	expected	actual	O-E	(O-E)^2	(O-E)^2/E			
Collection method								
Calls								
Settlement offer								
Debt collectors								

Legal notice								
Total	450							
					0.09	Chi square		
					0.13	Chi square(p=0.05) for df=3		
					1.54	Chi square (p=0.025) for df=3		
					1.98	Chi square (p=0.1) for df=3		

at 95% confidence interval there is a drift to 0.13 but this not strong enough to cause a significant change looking at the degree of freedom . so for risk tiering nothing will be done . this goes also with the chi square at 97.5% and 99%