Credit modelling

Consumer Financial Services

Partnered with client to validate the underwriting model and identified opportunities that could reduce write-offs

Validated the underwriting model by leveraging Logistic Regression and Random Forest algorithms to access the loan applications and predict the probability of write-offs

Financial statements’ reconciliation and analysis

Situation

- Company was trying to reduce its charge-off rates by validating their new underwriting model through a third-party assessment
- Partnered with the company to help assess loan applicants during the underwriting process before approving them for a loan

Accordion Value Add

- Leveraged Logistic Regression and Random Forest algorithms to predict the probability of write-off for a new loan and evaluated if advanced classification algorithms such as random forest could be used for developing risk model
- Used various techniques such as Boruta, step-wise regression to select important features for the model and used imputation techniques to deal with incomplete data
- Analyzed the delinquency data to help estimate proxy write-off for incomplete term loans
- Analyzed precision & recall values to select the best model across the techniques

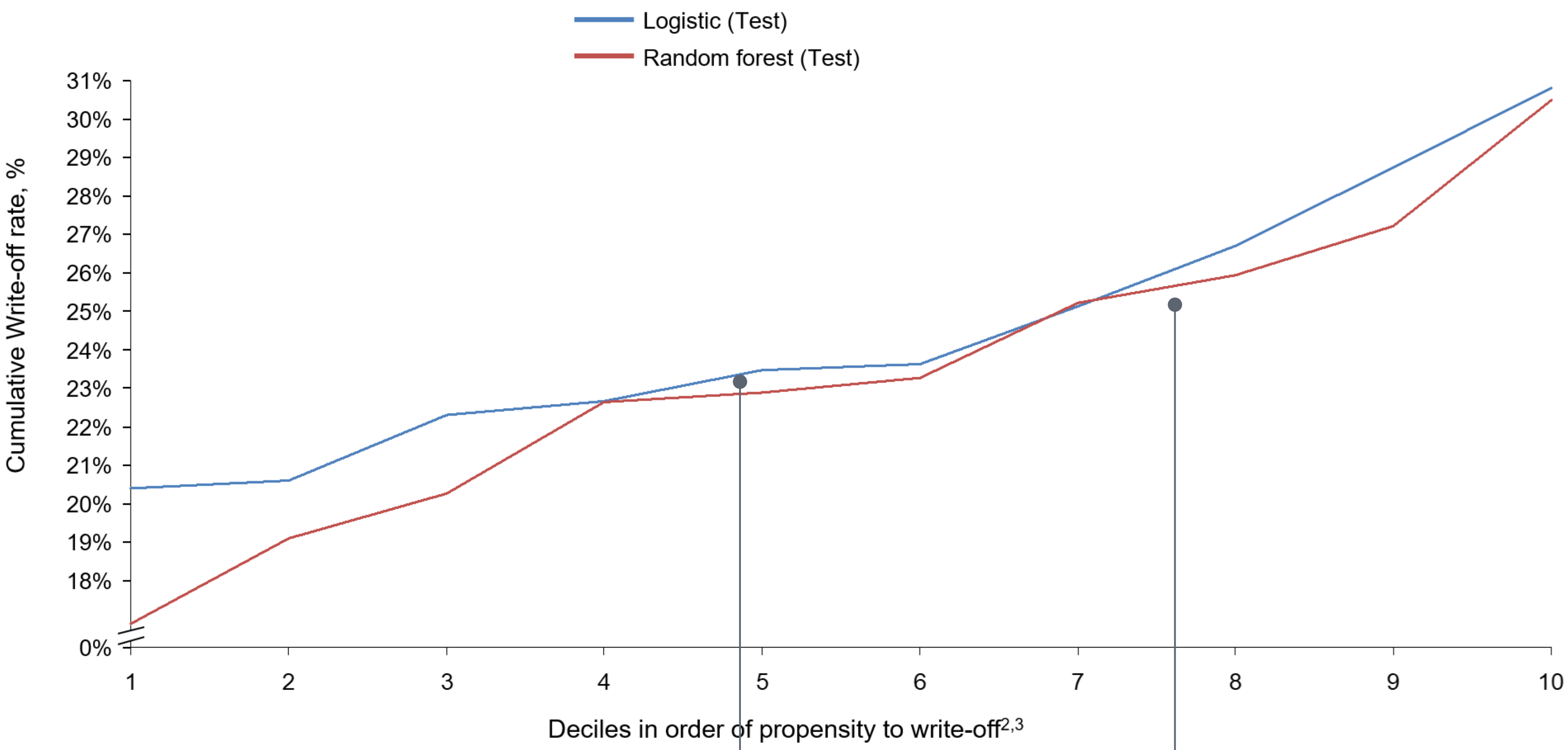
Impact

- The model developed to predict write-off rate for a loan applicant enabled the company to underwrite within their risk appetite
- Helped the client analyze various predictive algorithms to compare their respective impact over their current under-writing process; overall write-off rate was at least 30% lower using RF model or logistic model

Random forest sloped write-offs relatively better than logistic

Logistic & RF

Cumulative Write-off rates[1] by deciles, %



Analyzed expected write-off rates by deciles between algorithms

Features A, B and C played a critical role in predicting write-offs across iterations

Rank order of importance of independent variables, by Iterations[1,2]

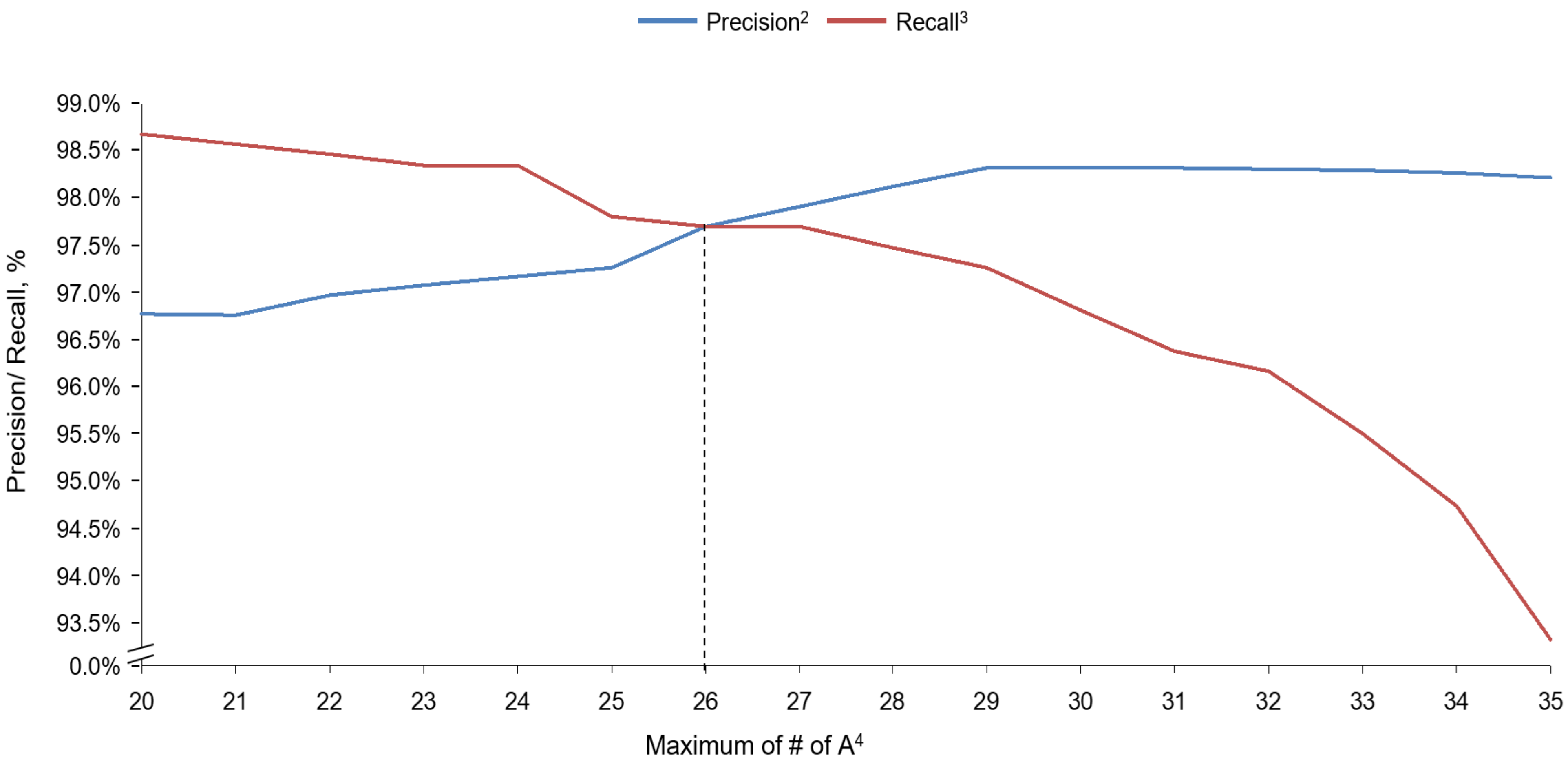
Variables	Iteration 1	Iteration 2	Iteration 4	Iteration 5	Iteration 7	Iteration 8
A	1	1	1	1	1	1
B	20	5	16	2	2	2
C	10	18	23	15	4	5
D	62	23	20	4	12	10
E		7	4	7	16	16
F	19	14		23	9	17
G	6			21	3	3
H		6	11		28	6
I	33		2	27	6	
J		30	30		8	8
K		12	10	14	49	
L		37	18	18	22	
M		39	14	24	51	
N		44	33	30	26	
O		9	3	3		
P		11	13	6		
Q		15	21	11		
R		33	22	12		
S				32	20	21
T		19	9		47	
U		64	6	5		
V		26		17		35
W		25	26		31	
X		29	17	38		
Y		4		25	57	

Ranked the features across e iterations to understand their importance

Maximum # of a days or more balanced the precision and recall of total write-off accounts

All Loans

Precision and Recall by maximum # of A[1]



Swap-in risk profile was better than swap-out risk across the important variables

Random forest

Risk Profile[1] by Swap-set category

Independent Variable	In-Ins	Swap-Ins	Swap-Outs	Out-Outs
A	11	8	14	16
B	737	780	714	718
C	647	695	614	613
D	175,640	376,745	99,378	91,603
E	654	690	637	618
F	6	8	4	5
G	305	343	279	302
H	98	153	54	72
I	8.5	11	6	7
J	722	726	744	565
K	134	186	103	116
L	65%	59%	68%	72%
M	164,614	260,589	160,300	160,300
N	5	4	6	5
O	12	17	10	10
P	0	0	-98[2]	0
Q	0	0	0	0
R	2	2	2	2