



*Association for
Financial Professionals®*

Predicting Membership Renewal in the Association for Financial Professionals

- Cytana Hubbard
- Galina Kozachenko
- Chris Klejdys
- Josh Miller
- Ayon Rahman





Association for
Financial Professionals®

GOAL

Goal of the project

Build a model that will allow AFP to predict members who are least likely to renew their membership into a new year.

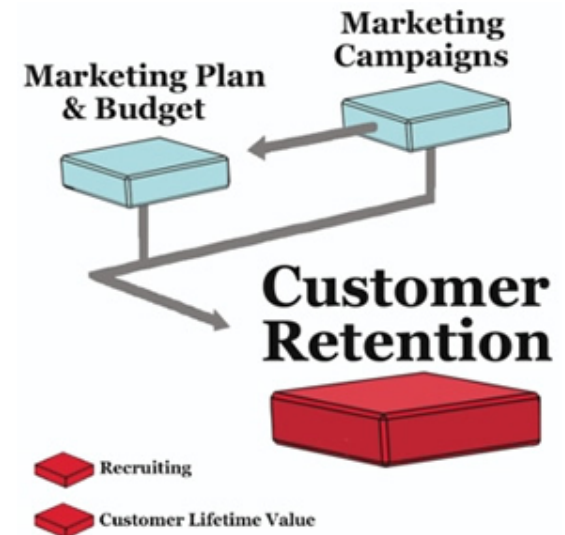
AFP Goals to the Board of Directors

Grow the corporate practitioner membership 5%

- Grew 0.47% in FY07
- Five consecutive years of growth

Achieve practitioner retention rate of over 80%

- 79.02% in FY06





EXPLORATION

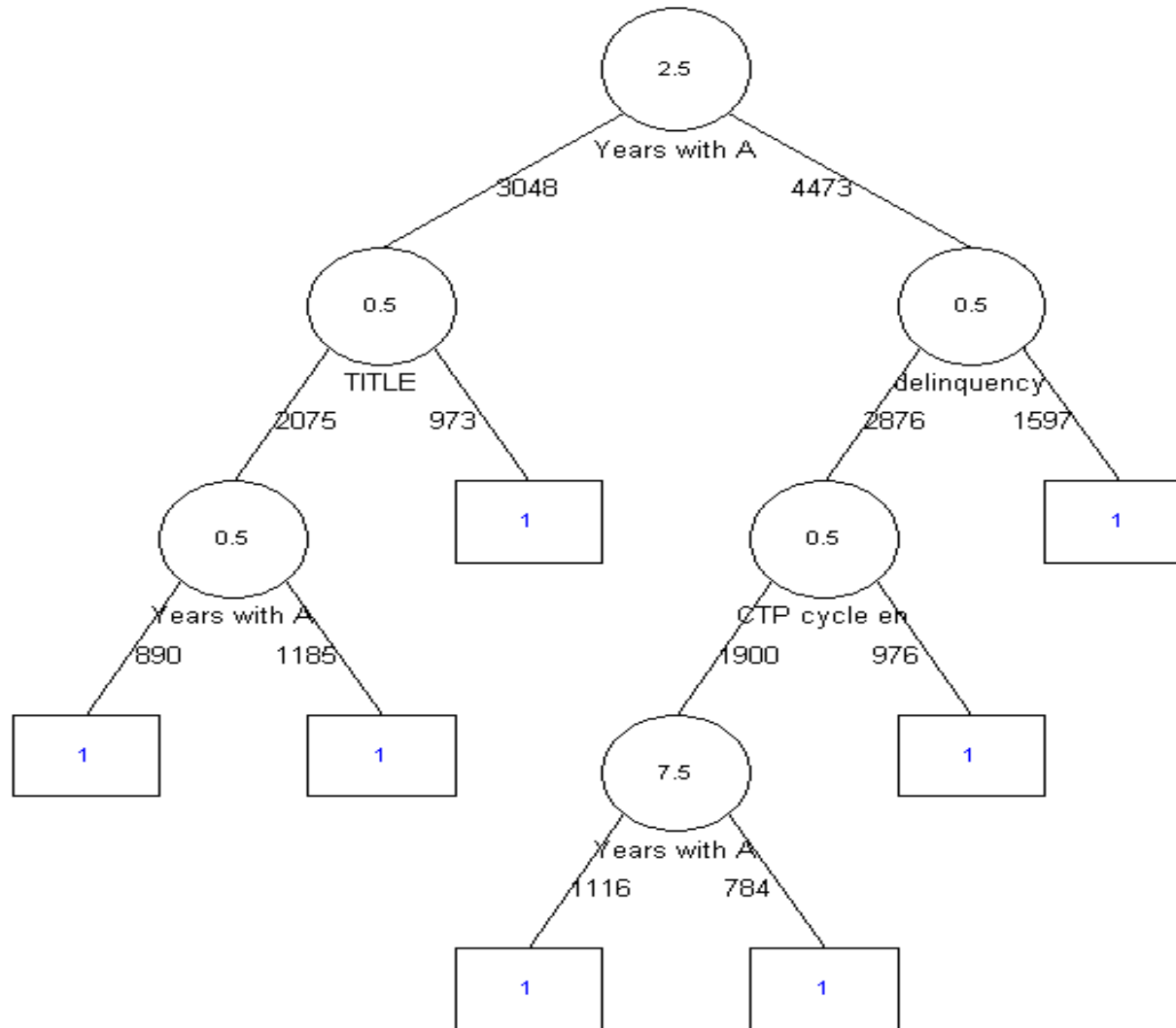
<u>Non-renews</u>			<u>Renews</u>		
<i>Summary measures for selected variables</i>			<i>Summary measures for selected variables</i>		
	Years with AFP	Years since delinquency		Years with AFP	Years since delinquency
Count	3116.000	3116.000	Count	11926.000	11926.000
Mean	3.932	0.505	Mean	5.809	0.720
Median	2.000	0.000	Median	5.000	0.000
Standard deviation	4.718	1.381	Standard deviation	5.280	12.145
Range	25.000	11.463	Range	27.000	12.145
21% non-renewal rate			79% renewal rate		
27% delinquent			22% delinquent		
6% no code			0.4% no code		
35% senior			40% senior		
57% corp practitioners			67% practitioners		
24% certified			49% certified		
9% in CTP cycle			20% in CTP cycle		
7% newsletter			24% newsletter		
8% discussion list			20% discussion list		
same participation in RT, PY and AN					



EXPLORATION

Practitioners	9429	63%			Associates	5796	39%	
non-renew	19%				non-renew	23%		
Summary measures for selected variables					Summary measures for selected variables			
	Years with AFP	Years since delinquency				Years with AFP	Years since delinquency	
Count	1775.000	1775.000			Count	1341.000	1341.000	
Mean	4.172	0.523			Mean	3.614	0.481	
Median	2.000	0.000			Median	1.000	0.000	
Standard deviation	4.822	1.442			Standard deviation	4.558	1.297	
Range	25.000	10.913			Range	25.000	11.463	
% delinquent	26%				% delinquent	28%		
no code	3%				no code	9%		
seniors	30%				seniors	41%		
blank/unemp ind	14%				blank/unemp ind	16%		
newsletter	7%				newsletter	7%		
discussion list	9%				discussion list	7%		
AN	15%				AN	25%		
CTP yes	7%				CTP yes	12%		
CTP no	14%				CTP no	17%		
certification	21%				certification	28%		

EXPLORATION

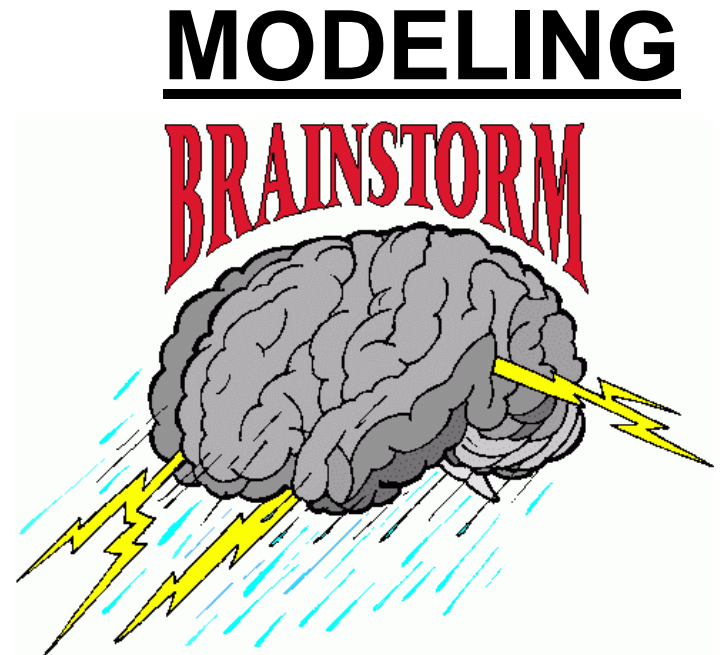


Models to run on 2 data sets: Associates
and Practitioners

- Logistic Regression
- Discriminant Analysis
- KNN
- Naïve Bayse

AFP Director of Research and Data
Standards: “It is more important to predict
non-renewals than renewals; the
misclassification ratio is 5 to 1 for
Practitioners and 4 to 1 for Associates”

Discriminant Analysis – best results



$$\left(\frac{b}{a+b}\right) \cdot p(C_1) \cdot q_1 + \left(\frac{c}{c+d}\right) \cdot p(C_2) \cdot q_2$$

Expected cost of misclassification

Logistic Regression for Associates

Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable) **0.5**

Classification Confusion Matrix			
		Predicted Class	
Actual Class		0	1
0		63	581
1		11	2241

Error Report			
Class	# Cases	# Errors	% Error
0	644	581	90.22
1	2252	11	0.49
Overall	2896	592	20.44

cost renew 1
 cost non-renew 4
 prob non-renew 0.22
0.797723

	Overall Accuracy	Sensitivity	Specificity	Expected Cost
	0.79558011	0.097826087	0.004884547	0.79772299
0	0.222375691	1	1	0.78
0.05	0.265883978	0.99068323	0.941385435	0.742479397
0.1	0.368093923	0.948757764	0.797957371	0.667499917
0.15	0.511740331	0.822981366	0.577264654	0.606042827
0.2	0.616367403	0.703416149	0.408525755	0.579643878
0.25	0.662983425	0.647515528	0.33259325	0.569609071
0.3	0.71961326	0.49689441	0.21669627	0.61175601
0.35	0.787638122	0.183229814	0.039520426	0.749583696

Models for Practitioners

LR	output 4	Residual df	4614
		Residual Dev.	4010.4009
		% Success in training data	19.679931
		# Iterations used	7
The Regression Model		Multiple R-squared	0.1255866

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-0.6869095	0.0690726	0	*
Years with AFP	-0.0698475	0.0102078	0	0.932536
BlankUnemployed	0.5552537	0.1255593	9.77E-06	1.7423831
delinquency	0.7181777	0.1031927	0	2.0506928
NO CODE	2.5091314	0.3773286	0	12.294249
TITLE	-0.1918304	0.0867243	0.0269697	0.8254468
NEWSLETTER	-1.1197598	0.1425239	0	0.3263582
DISCUSSION	-0.5690045	0.134673	2.389E-05	0.5660887
CTP cycle end YES	-1.1338776	0.1461905	0	0.3217831
CTP cycle end NO	-0.8556523	0.1100288	0	0.4250059

Test Data scoring - Summary Report

According to relative occurrences in training data			
Class	Prob.		
0	0.1967993	<-- Success Class	
1	0.8032007		
Cut off Prob.Val. for Success (Updatable)		0.2	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	0	1	
0	242	100	
1	556	952	
Error Report			
Class	# Cases	# Errors	% Error
0	342	100	29.24
1	1508	556	36.87
Overall	1850	656	35.46

Models for Associates

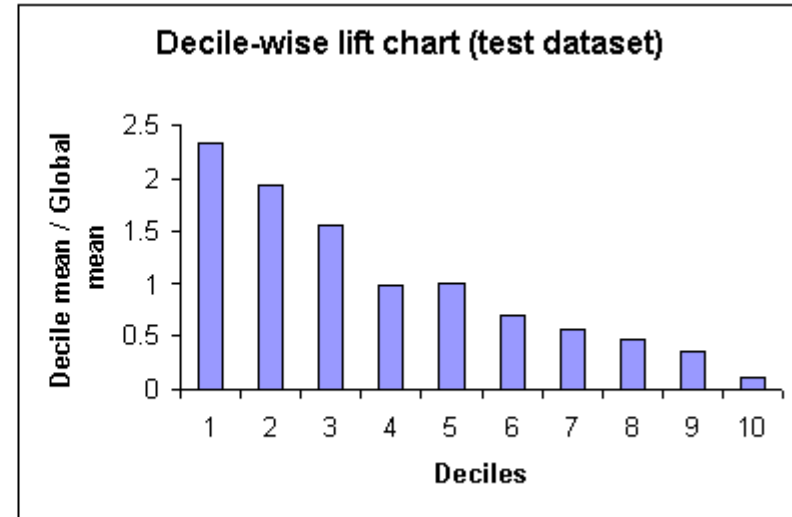
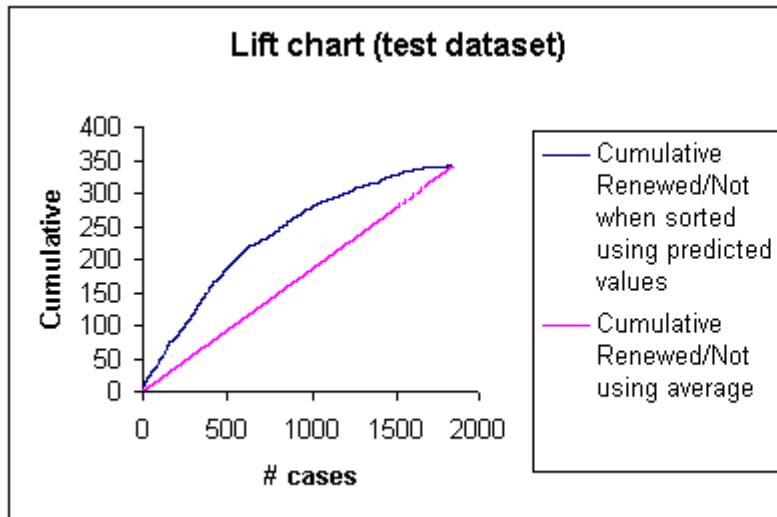
LR		Residual df	2887
		Residual Dev.	2733.2563
		% Success in training data	22.237569
		# Iterations used	7
The Regression Model		Multiple R-squared	0.1094482

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-0.6505982	0.0796922	0	*
years with AFP norm	-1.6643167	0.3292501	4.3E-07	0.18932
delinquency	0.5541642	0.1209827	4.64E-06	1.7404858
NO CODE	2.7890627	0.3219627	0	16.265766
TITLE	-0.1805019	0.0994518	0.0695289	0.8348512
NEWSLETTER	-0.9703825	0.1781178	5E-08	0.3789381
DISCUSSION	-0.3591167	0.181672	0.048072	0.6982929
CTP cycle end YES	-0.7554134	0.1452486	0.0000002	0.4698163
CTP cycle end NO	-0.8532068	0.1262424	0	0.4260465

Test Data scoring - Summary Report

According to relative occurrences in training data			
Class	Prob.		
0	0.2223757	<-- Success Class	
1	0.7776243		
Cut off Prob.Val. for		0.25	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	0	1	
0	163	105	
1	296	595	
Error Report			
Class	# Cases	# Errors	% Error
0	268	105	39.18
1	891	296	33.22
Overall	1159	401	34.60

Logistic Regression for Practitioners





RECOMMENDATIONS

Corporate Practitioners

	Profit	Upfront cost	ROI	Sensitivity	Specificity	False Pos	False Neg	Overall	Exp. Cost
LR	\$26,448.00	\$28,728.00	92%	71%	63%	70%	10%	35.46%	1.21
DA	\$26,448.00	\$28,044.00	94%	70%	64%	69%	10%	34.76%	1.21
KNN	\$25,872.00	\$29,304.00	88%	71%	62%	70%	10%	36.32%	1.20
NNB	\$26,292.00	\$34,812.00	76%	78%	54%	72%	8%	41.78%	1.21

Associates

	profit	upfront cost	ROI	Sensitivity	Specificity	False Positive Error	False neg	Overall	Exp Cost
LR	\$20,640.00	\$16,524.00	125%	61%	67%	64%	15%	34.60%	1.06
DA	\$20,808.00	\$16,128.00	129%	60%	68%	64%	15%	33.82%	1.06
KNN	\$22,032.00	\$21,060.00	105%	71%	56%	68%	14%	40.98%	1.05
NNB	\$23,616.00	\$21,528.00	110%	74%	55%	67%	12%	40.55%	1.08

- Go with Logistic Regression if the goal is to minimize the upfront cost and maximize the ROI
- Go with the Naïve Bayse if the goal is to increase the number of renewals
- When the other two variables become available after the re-design of the web site and web store, introduce two new variables into the models to observe whether the predictive ability increases.