BOIRE FILLER
G·R·O·U·P

Insurance Pricing Models Using Predictive Analytics

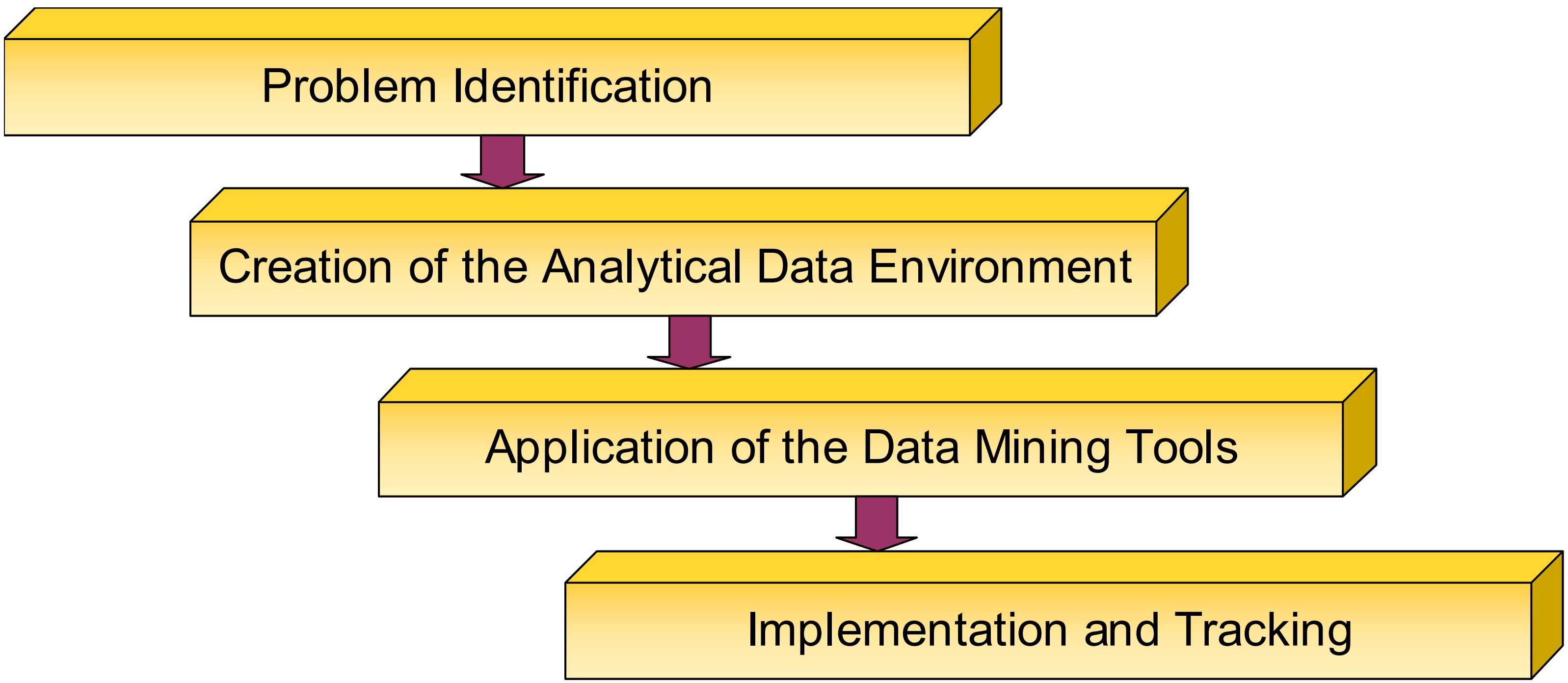
March 5th, 2012

Agenda

- Background/Information
- Why Predictive Analytics is Required
- Case Study-AMA
- An Effective Decision Tool for front-line Underwriting and Brokers
- Conclusion

Background/Information

- The standard 4 step analytical process is required for solutions to be successful in any predictive analytics exercise within any business:



Why Predictive Analytics is Required

- What does the 4 step approach mean within the context of building insurance pricing solutions?

- 1st Step
  - What is the real problem here?

- Insurance companies have always been able to provide pricing solutions using data
  - The actuarial discipline is comprised of individuals with strong quantitative and mathematical skills
  - This discipline is not new and is considered the analytical engine of most insurance companies

- So why do they need data miners/predictive analytics practitioners?

Why Predictive Analytics is Required

- Historically, the use of analytics by actuaries has been about comparing groups
  - Differentials are then used to create a price or premium for policy holder
  - Example: What premium should a male driver pay for automobile insurance? What about a female?

Gender	# of policy holders	Total Claim Loss	Avg. Loss Cost	Differential
Male	100,000	$50,000,000	$500	**1.22**
Female	80,000	$24,000,000	$300	**0.73**
Total	180,000	$74,000,000	$411	**1.00**

•In this simplistic example, assuming gender is statistically significant, males and females would be charged a base premium multiplied by the differential:

Male: **1.22** X base premium of $600= **$732**

Female: **.73** X base premium of $600= **$534**

Why Predictive Analytics is Required

- Now, contribute more factors: gender, age, and distance to work:

	Under 30 Kilmetres to work	Under 30 Kilmetres to work	Over 30 Kilmetres to work	Over 30 Kilmetres to work	Total
	under 25 yrs	over 25 yrs	under 25 years	over 25 years	
Male	1.16	1.09	1.95	1.70	1.22
Female	0.61	0.49	0.97	0.91	0.73
total	1.16	1.22	0.88	1.03	1.00

Male under 25 years old who drives more than 30 kilometres to work would be charged: $600 X 1.95 = **$1170**

Female over 25 years old who drives under 30 kilometres to work would be charged: $600 X .49 = **$294**

So why isn't this sufficient for pricing purposes?

Why Predictive Analytics is Required

Groups	# of Records	Differential
Male Over 25 years and drives over 30 kilometres to work	100,000	1.7
Total # of policies	300,000	1

- Using the technique of calculating risk based on group differentials, we have 100,000 or 1/3 of the entire portfolio which will obtain the same level of risk.
  - Is it possible to get more granular in calculating risk for smaller groups of records?

- Another drawback to using the above method is multicollinearity or interaction between variables
- These kind of limitations are overcome through the use of MVA (Multivariate Analysis) or predictive analytics type solutions:
  - Outcome is a score for each individual
  - Solution takes into account the interaction between variables

Why Predictive Analytics is Required

Two Reasons why Predictive Analytics has not been used historically:

1. Technology	2. Analytics Practitioners
• Inability to employ analytics at an individual policy level	• Lack of knowledge and expertise in data manipulation techniques to create analytical file at individual policy record

Does this mean that traditional Actuarial techniques don't work?

Why Predictive Analytics is Required

Deciles using Premium as the solution	AVG. Claim Amount	% OF TOTAL Claim amounts in Interval	Claim/Premium Ratio
1	$483.75	17%	0.74
2	$467.44	16%	0.78
3	$293.40	10%	0.53
4	$363.71	13%	0.81
5	$343.05	12%	0.86
6	$238.43	8%	0.68
7	$194.57	7%	0.65
8	$219.14	8%	0.88
9	$83.88	3%	0.42
10	$160.03	6%	1.07

Premium as determined by traditional actuarial approaches works quite well in assessing claim risk(avg.claim amount).

In this example,though,the tool is unable to target policies by claim/premium ratio.

Deciles using Predicted Analytics as the solution	Avg. Claim Amount	% OF TOTAL Claim Amounts in Interval	Claim/Premium Ratio
1	981	25.00%	0.88
2	643	17.00%	0.77
3	546	14.50%	0.75
4	419	12.50%	0.66
5	329	8.50%	0.59
6	297	7.50%	0.61
7	250	5.50%	0.59
8	168	4.50%	0.47
9	129	3.50%	0.45
10	56	1.50%	0.26

Here is gains/decile chart using predictive analytics solutions.

Notice the increased rank ordering capability both on observed avg. claim amount and claim/premium ratio

Case Study - AMA

- AMA issues insurance policies for property homeowners
- Challenges:
  - Loss Ratios have been steadily increasing over last few years
  - Demonstrate how predictive analytics solutions can improve upon their existing methods of assigning risk to homeowners

- What will be our ‘ACE IN THE HOLE’ here?
- The DATA
  - Create the dependent variable of loss cost (claim frequency X severity)
  - Create independent variables that can potentially predict loss cost
  - Create an analytical file where each record represents one policyholder with hundreds of different fields of information

***ULTIMATELY INCREASED DATA GRANULARITY**

Case Study - AMA

- Sources of data used to create the analytical files are:

1. Prop_Claim_tran file • 90,490 records	2. Prop_Prop_tran file • 714,594 records	3. Prop_Data_Loc_ Postalcodes • 488,980 records

- This source data was used to create our analytical file of 60M records

- First step: perform data audit on all source data
  - Data diagnostic reports and frequency distributions to assess data quality and data integrity

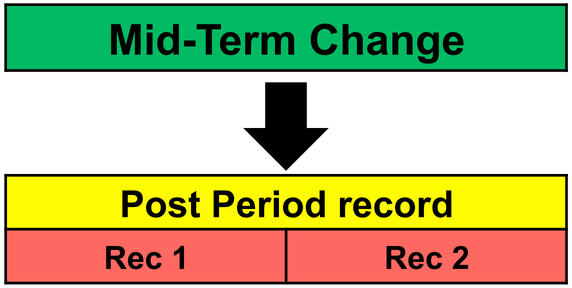
- Data audit results provide key insights into what variables should be created for this exercise.

Case Study-AMA

- Unique Business and Data Challenges
  - Working with stakeholders with strong quants knowledge
  - Determining the appropriate pre post window
    - Do we have enough claims in post period to build model?
    - How recent are they?

Pre Period	**Post Period**
Independent Variables	Dependent Variables of Claim and Amount

- Accounting for mid-term changes in term is unique in creating insurance risk analytics solutions



Case Study - AMA

- Key variables in the analytical file were created:
  - Prior claim history
  - Territory/geography
  - Policy statistics/coverage type
  - Location information
  - Stats Can Data
  - Etc.

- Correlation and EDA reports were run as part of the model building exercise

Case Study - AMA

- Sample Correlation

Variable	Correlation	P-Value
Rate Territory=M1	-0.0218	0.0002
% Clerical Occupations	-0.0201	0.0005
Rider=RVC	0.0199	0.0006
gross limit/liability limit	0.0196	0.0007
Rate Territory=3	0.0194	0.0008
%intraprovincial migrants	0.0181	0.0018

- Sample EDA

Variable	Count	Range	Observed Claim Amount
FSA-CHAID Variable	**29800**	**Average**	**$774.39**
	4361	1	$1,498.00
	8766	2	$860.00
	3467	3	$787.00
	5455	4 to 5	$592.00
	7751	6 to 7	$393.00
Number of Riders	**29800**	**Average**	**$774.39**
	23061	0	$674.67
	6739	1 to 7	$1,115.66

Case Study - AMA

- Most variables in the model were consistent with conventional property insurance underwriting concerns:
  - Location of property - FSA (based on first 3 digits of postal code)
  - Coverage Limits
  - Previous claim history
  - Policy Riders
  - Statistics Canada (socio-demographic)
    - e.g. education and occupation
  - Client Age

- The next step was to validate this model and observe its actual performance

Case Study - AMA

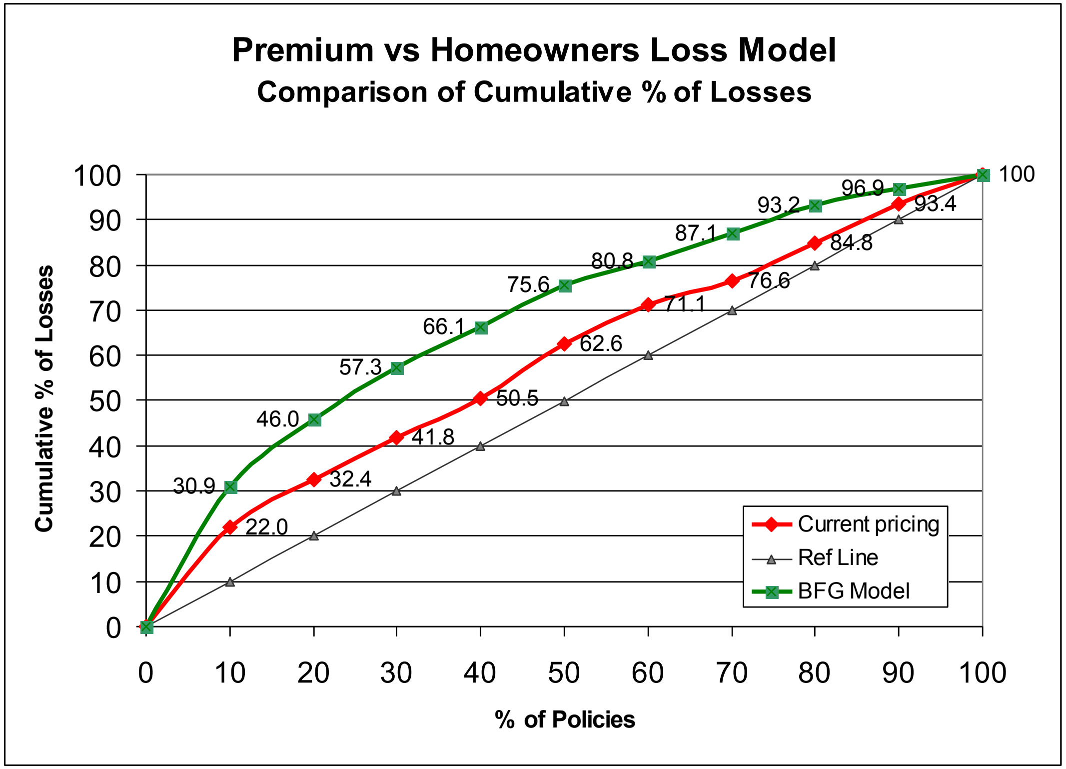
- Decile/Gains Chart

Deciles ranked by predictive analytics model	Average Premium	Average Actual Loss Amount	Claim/Premium Ratio
1	1354	1895	140
2	1045	1240	119
3	907	740	82
4	865	798	92
5	791	935	118
6	746	800	107
7	721	553	77
8	691	430	62
9	666	484	73
10	608	176	29

Premium is still an effective tool but the predictive analytics solution is superior

Case Study- AMA

Model Performance: "Lift" vs. Current Rating Method



- The Line Chart depicts the cum. percentage of actual losses in the portfolio by the model score(green l,ine) and the current premium(red line) being charged by the company
- The distance between the line representing losses predicted by the model vs. current pricing structure represents the "lift", or increased accuracy in loss prediction provided by the model
- In the top 20% of risks, the model captures 46% of losses compared to 32.4% for premium

BOIRE FILLER G·R·O·U·P

An Effective Decision Tool for front-line Underwriting and Brokers

Request for Insurance quotation → Acquisition of Scoring Data → Calculation of Score → Delivery of Score

Sales and underwriting processes designed to optimize underwriting performance

Group 4 - Top Priority Client
- "Quick Quote"
- Best sales effort

Sales Objective: 100%
U/W Action: None

Group 3 – Priority Client
- Quote and collect info from customer
- Bind coverage

Sales Objective: 60%
U/W Action: Review

Group 2 - Low Priority Client
- Thorough underwriting review
- Order reports and bind coverage once info verified

Sales Objective: 30%
U/W Action: Detailed analysis

Group 1 - Do Not Sell
- Customer to provide detailed documentation for underwriting

Sales Objective: 0%
U/W Action: Maximize premium & get off risk

Conclusions

- Current tracking of model indicates model is performing as expected
- Currently developing auto claim risk models for same client
- Most of work in insurance claim risk is in the auto sector
- More competition now in producing these models
- Predictive Analytics practitioners have the edge
  - **WORKING THE DATA**