Institute and Faculty of Actuaries

Predictive modelling around the world

Peter Banthorpe, RGA
Kevin Manning, Milliman

11 November 2013

Agenda

- Introduction to predictive analytics
- Applications overview
- Case studies
- Conclusions and Q&A

Institute and Faculty of Actuaries

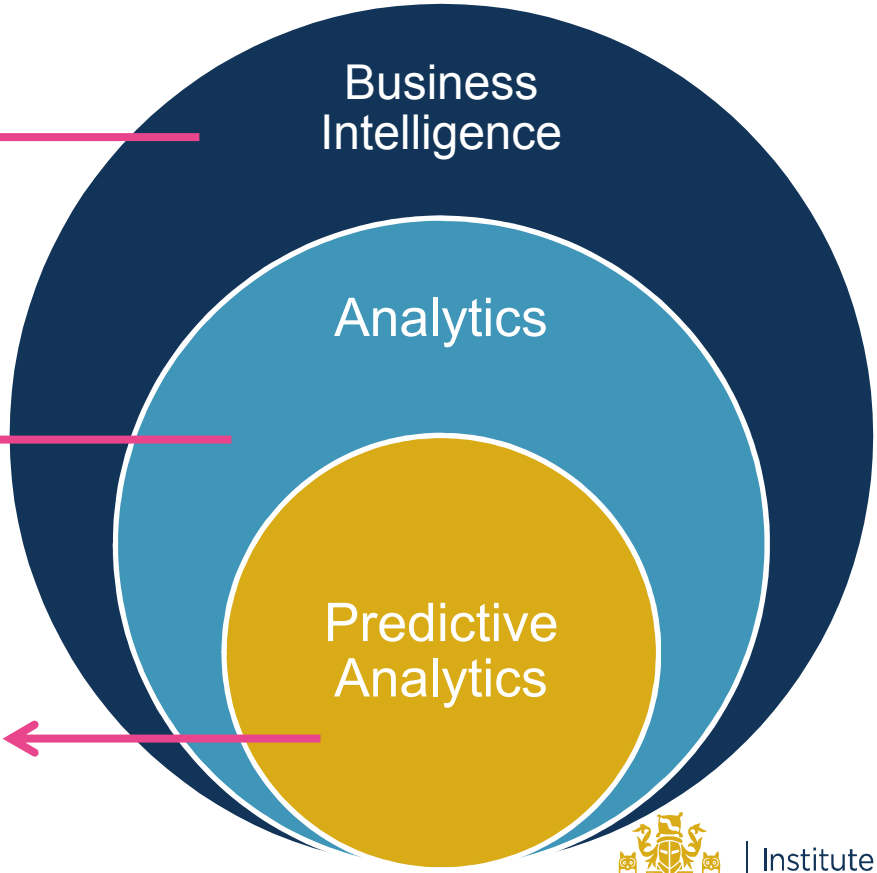
Introduction to predictive analytics

What are predictive analytics?

Tools and technologies for analysing and understanding business performance

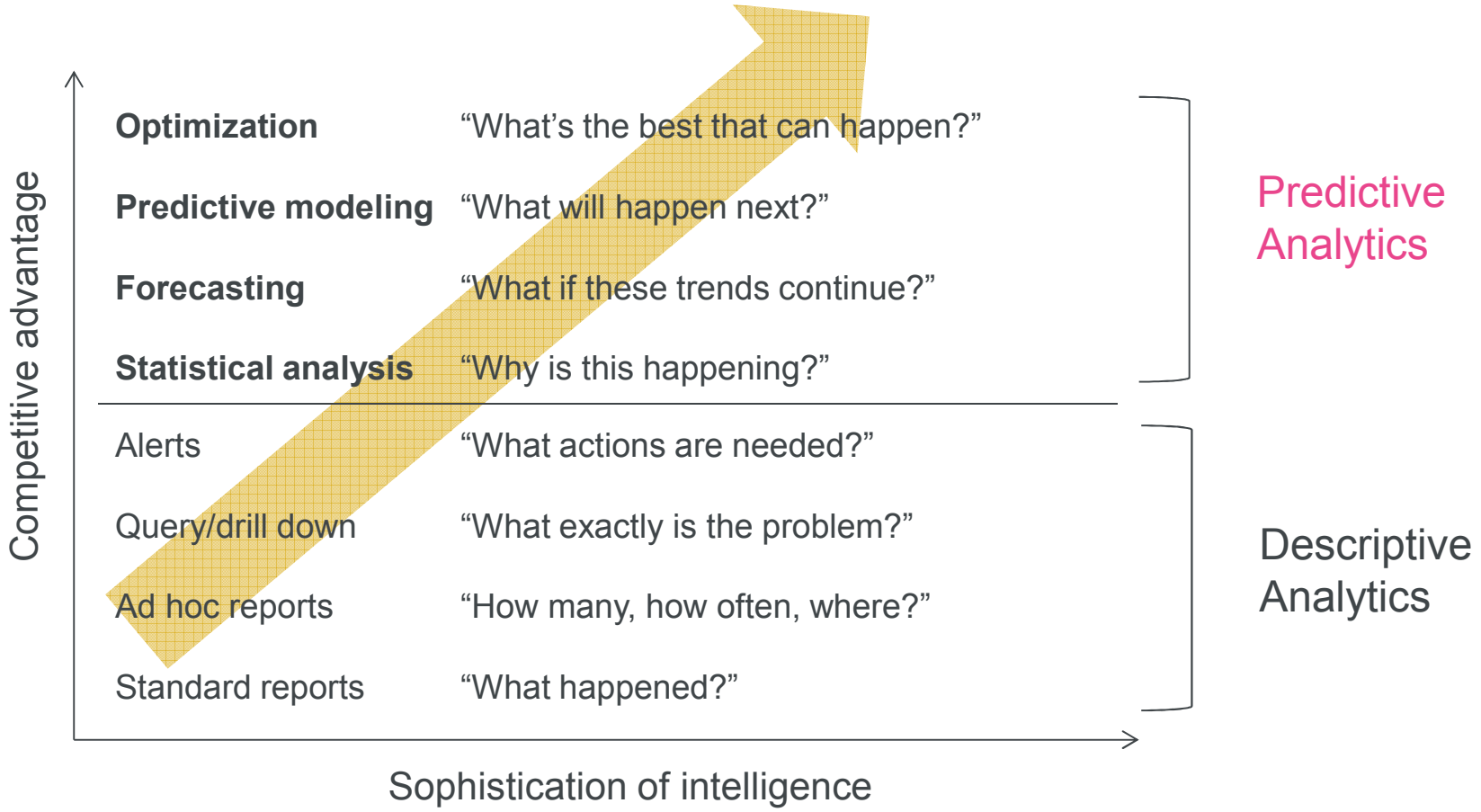
Extensive use of data with statistical and quantitative analysis

A process by which current or historical facts are used to create predictions about future events or behaviour.



Institute and Faculty of Actuaries

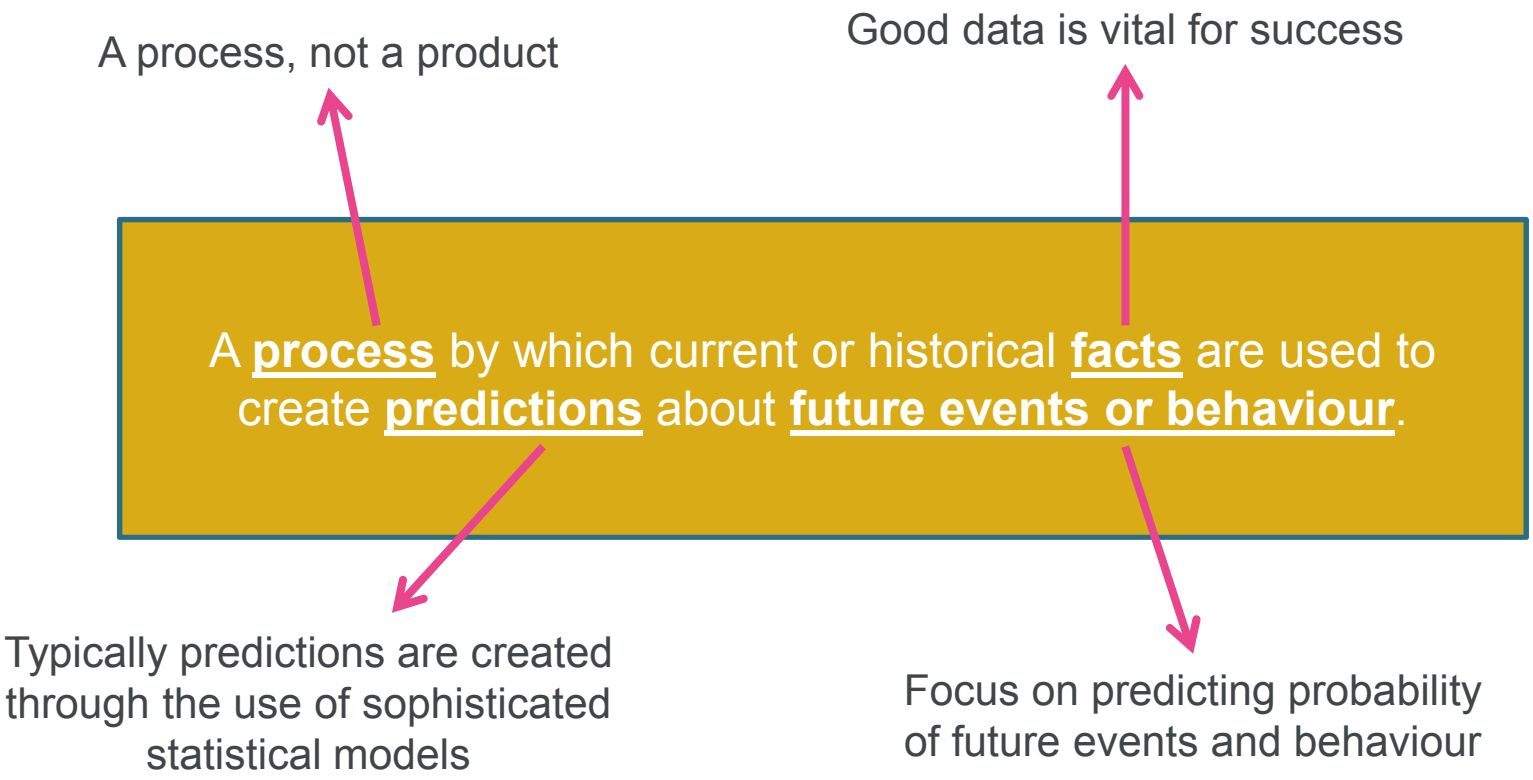
Descriptive versus Predictive Analytics



Source: Competing on Analytics: The New Science of Winning by Thomas Davenport and Jeanne G. Harris (Harvard Business School Press, 2007)

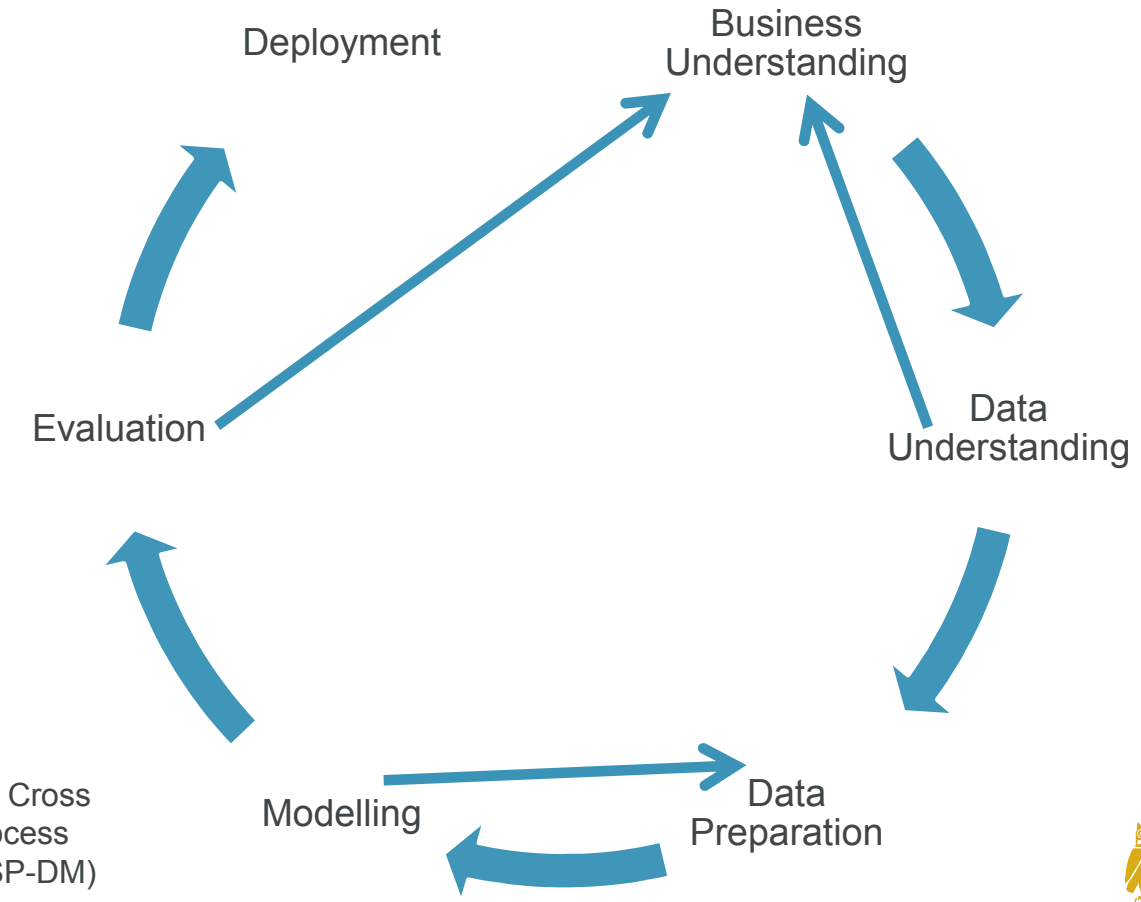
Institute and Faculty of Actuaries

What are predictive analytics?



Institute and Faculty of Actuaries

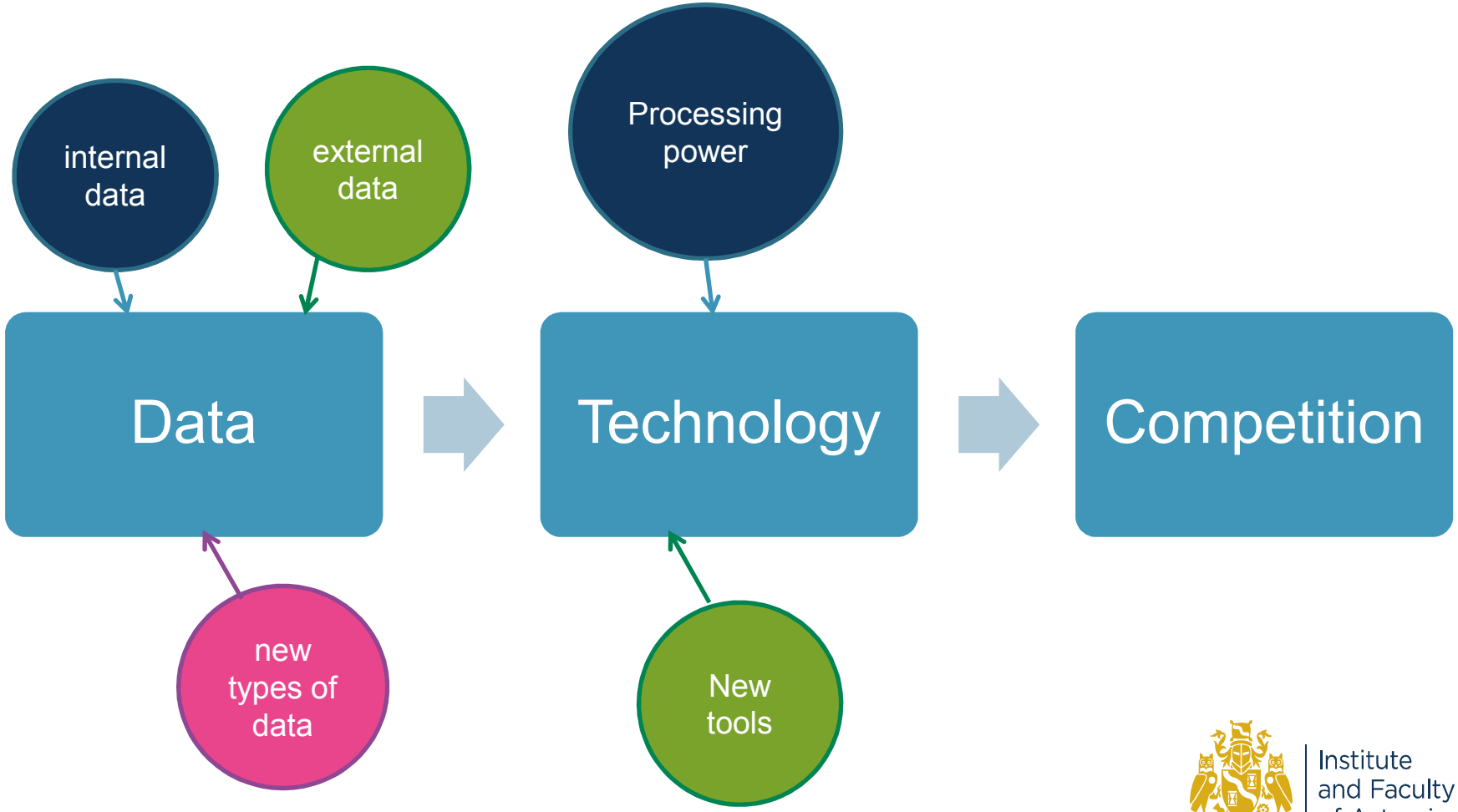
Predictive modelling process



Source: adapted from Cross Industry Standard Process for Data Mining (CRISP-DM)

Institute and Faculty of Actuaries

Why predictive analytics?



Institute and Faculty of Actuaries

Non-life example (USA)

Profitability & Retention

High-profit, low-retention customer segment in red

	Retention								Exposure
Loss Ratio	98.5%	97.7%	96.7%	95.8%	94.5%	91.5%	87.3%	80.8%	Total
35.0%	3.0%	1.6%	1.8%	0.9%	0.4%	0.1%	0.4%	0.5%	9%
42.7%	3.6%	2.2%	1.8%	0.8%	0.7%	0.3%	0.7%	0.6%	11%
45.8%	2.0%	3.0%	3.5%	2.4%	1.6%	1.1%	1.2%	0.5%	15%
49.5%	0.7%	2.0%	3.6%	3.4%	2.6%	1.9%	1.2%	0.6%	16%
55.5%	0.3%	1.0%	2.6%	3.4%	3.7%	2.7%	1.5%	0.7%	16%
58.2%	0.0%	0.4%	1.5%	2.7%	3.6%	4.2%	2.0%	0.8%	15%
61.4%	0.0%	0.1%	0.5%	1.4%	2.0%	2.9%	2.0%	1.0%	10%
75.4%	0.0%	0.0%	0.1%	0.7%	1.2%	2.7%	1.5%	2.0%	8%
Exposure Total	10%	10%	15%	16%	16%	16%	11%	7%	100%

Source: EagleEye Analytics

Low-profit, high-retention customer segment in red

	Retention								Exposure
Loss Ratio	98.5%	97.7%	96.7%	95.8%	94.5%	91.5%	87.3%	80.8%	Total
35.0%	3.0%	1.6%	1.8%	0.9%	0.4%	0.1%	0.4%	0.5%	9%
42.7%	3.6%	2.2%	1.8%	0.8%	0.7%	0.3%	0.7%	0.6%	11%
45.8%	2.0%	3.0%	3.5%	2.4%	1.6%	1.1%	1.2%	0.5%	15%
49.5%	0.7%	2.0%	3.6%	3.4%	2.6%	1.9%	1.2%	0.6%	16%
55.5%	0.3%	1.0%	2.6%	3.4%	3.7%	2.7%	1.5%	0.7%	16%
58.2%	0.0%	0.4%	1.5%	2.7%	3.6%	4.2%	2.0%	0.8%	15%
61.4%	0.0%	0.1%	0.5%	1.4%	2.0%	2.9%	2.0%	1.0%	10%
75.4%	0.0%	0.0%	0.1%	0.7%	1.2%	2.7%	1.5%	2.0%	8%
Exposure Total	10%	10%	15%	16%	16%	16%	11%	7%	100%

Source: EagleEye Analytics

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Applications Overview

Predictive Analytics Projects Worldwide*

* Which the speaker knows about!

UK:
- Basis Setting (mortality, morbidity and lapses)
- Postcode pricing model
- Enhanced experience analysis
- Predictive underwriting on credit rating agency and bank data
- Broker Quality

Europe:
- Predictive underwriting on bancassurance data

South Africa:
- Enhanced Experience Analysis
- Predictive underwriting on bank and credit card data

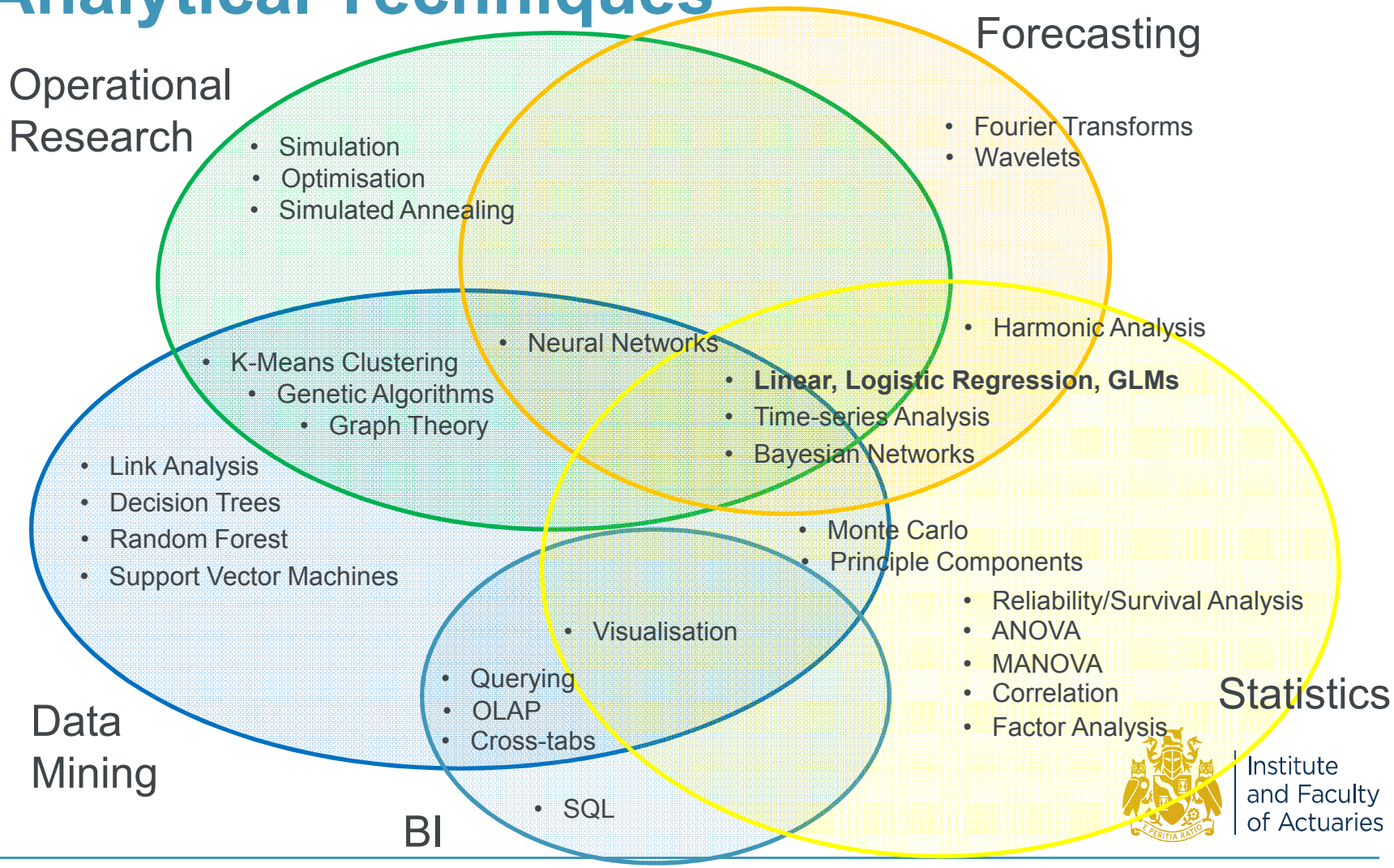
India:
- Claims Fraud Prediction

Australia:
- Predictive underwriting / cross sell on bancassurance data

USA:
- Pricing override model for group LT disability
- Lapse basis
- Predictive underwriting on Non-Life data
- Term Tail Lapses
- Mortality prediction on credit rating agency data

Asia:
- Predictive underwriting on bancassurance data
- Finer price segmentation
- Propensity to buy
- Cross sell of insurance on bank data

Predictive Analytics Covers Many Analytical Techniques



Institute and Faculty of Actuaries

Modified from a version presented by John Elder, www.datamininglab.com, 2012.

Key Themes Worldwide

- Availability of data is a key hurdle
  - Sufficient volume of data linking both predictors and outcome is needed
  - Ability to access useful data in different areas of a company
  - Data protection concerns
- Predictive underwriting is most popular application
  - Very popular with banks, especially in Asia….
  - BUT linking bank data to underwriting data is a common hurdle
- Generalised Linear Models being widely used
  - Predictive power sufficient
  - Relatively straight forward to understand / explain

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Case Studies

Case Studies for Today

1. Predictive Underwriting model

- Data: Bank client data, underwriting outcomes
- Model: GLM
- Decision: Who to offer simplified underwriting to

2. Pricing override model

- Data: Rating factors and profitability metrics
- Model: Classification and Regression Trees
- Decision: Where to override the pricing manual

Institute and Faculty of Actuaries

Case Study 1: Predictive Underwriting Model

- Client: Bancassurer in Asia with large customer pool, but low penetration in life product
- Goal: to predict UW decisions on its existing customers
- Major challenges - very limited data
  - A total of about 8k-9k full UW cases
  - Target variable UW decision, with very low declined/rated cases, ~3.0%
  - Many missing values due to old time, especially for sub-STD
  - Not all information collected at the time of UW

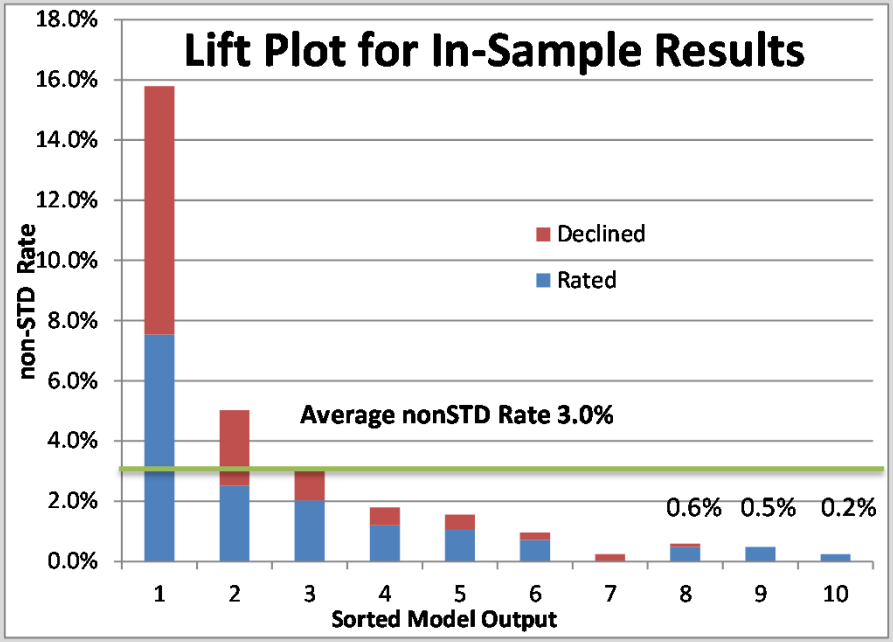
Modeling Approach / Key Variables

- GLM with binomial and logistic link function
- About a dozen of predictor variables that are statistically significant for prediction & readily available in client database
- Key predictor variables
  - "Positive" means the probability of being a standard rate case increases if the value goes up; otherwise, it is "Negative"

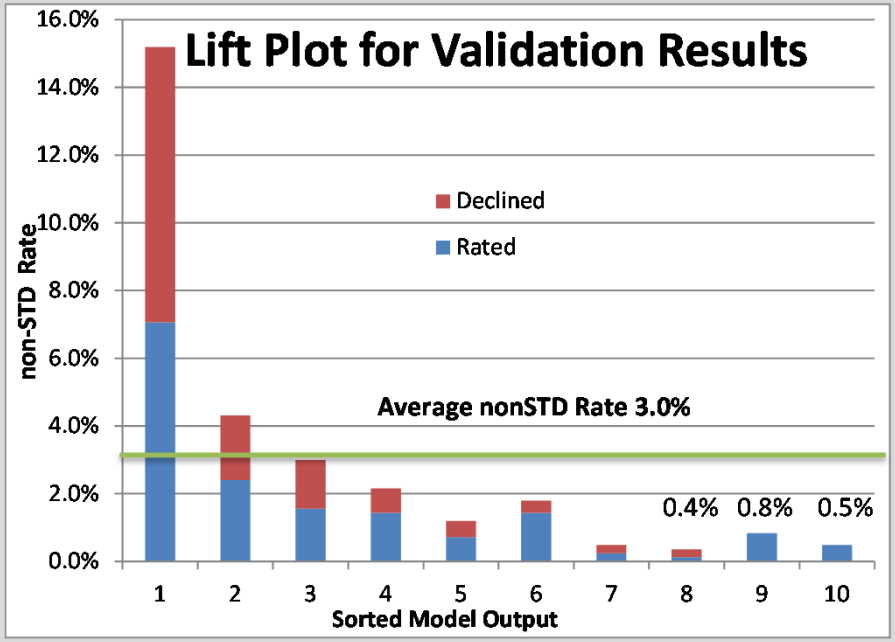
Name	Type	Note
Age_At_Entry	Numeric	Negative; less likely to qualify for STD as age goes up
Branch	Categorical	Proxy of geographic locations
Asset Under Management	Numeric	Positive; more likely to qualify for STD with large AUM
Customer_Segment	Categorical	Positive for "Gold", negative for other
Nationality	Categorical	Positive for domestic; negative for certain others

Model Results – Lift Plots

- In-sample results show model performance under optimal condition
- May over-fit data
- 0.5% of sub-STD in top 30%

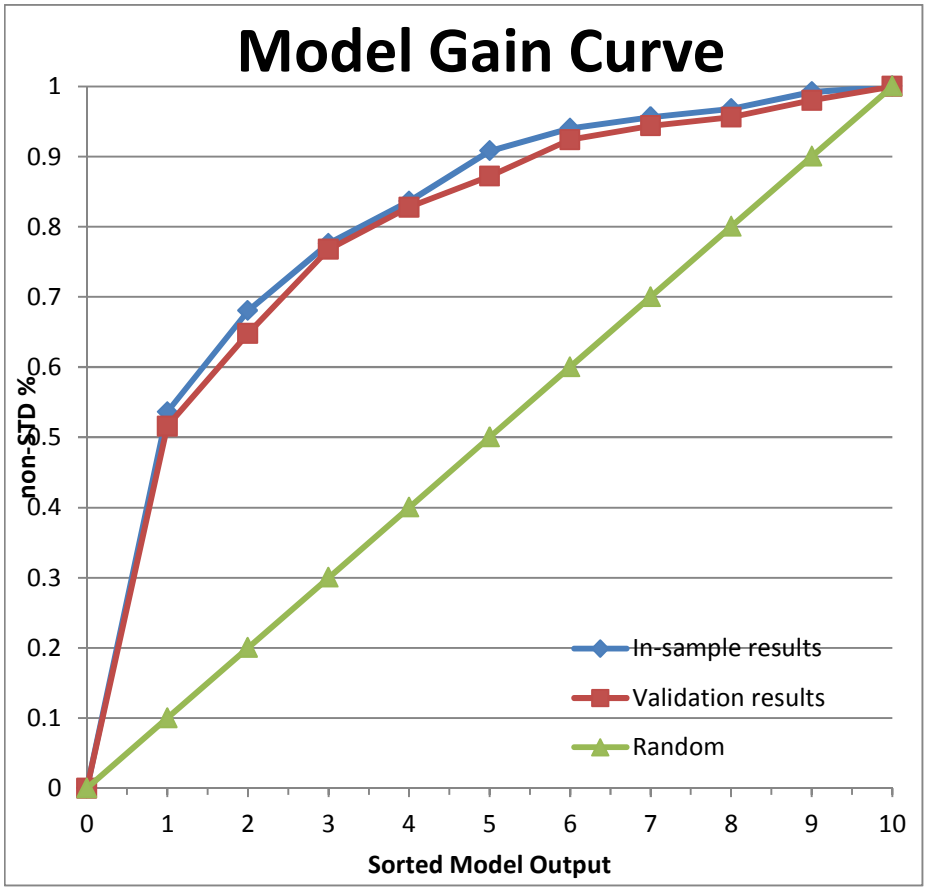


- Validation results are a better test of model performance in real business
- 0.6% sub-STD in the top 30%



Model Results – Gain Curve

- Another way to understand model capability to differentiate STD from sub-STD
  - Best 30% of model outputs contains about 5% of total non-STD
  - Lowest 30% captures about 75% of bad risks



Institute and Faculty of Actuaries

Classification and Regression Tree (CART) Model

- Both classification and regression
- Non-parametric approach (no insight in data structure)
- CART tree is generated by repeated partitioning of data set
  - Data is split into two partitions (binary partition)
  - Partitions can also be split into sub-partitions (recursive)
- Results are very intuitive
  - Identify specific groups that deviate in target variable
  - Yet, algorithm is very sophisticated

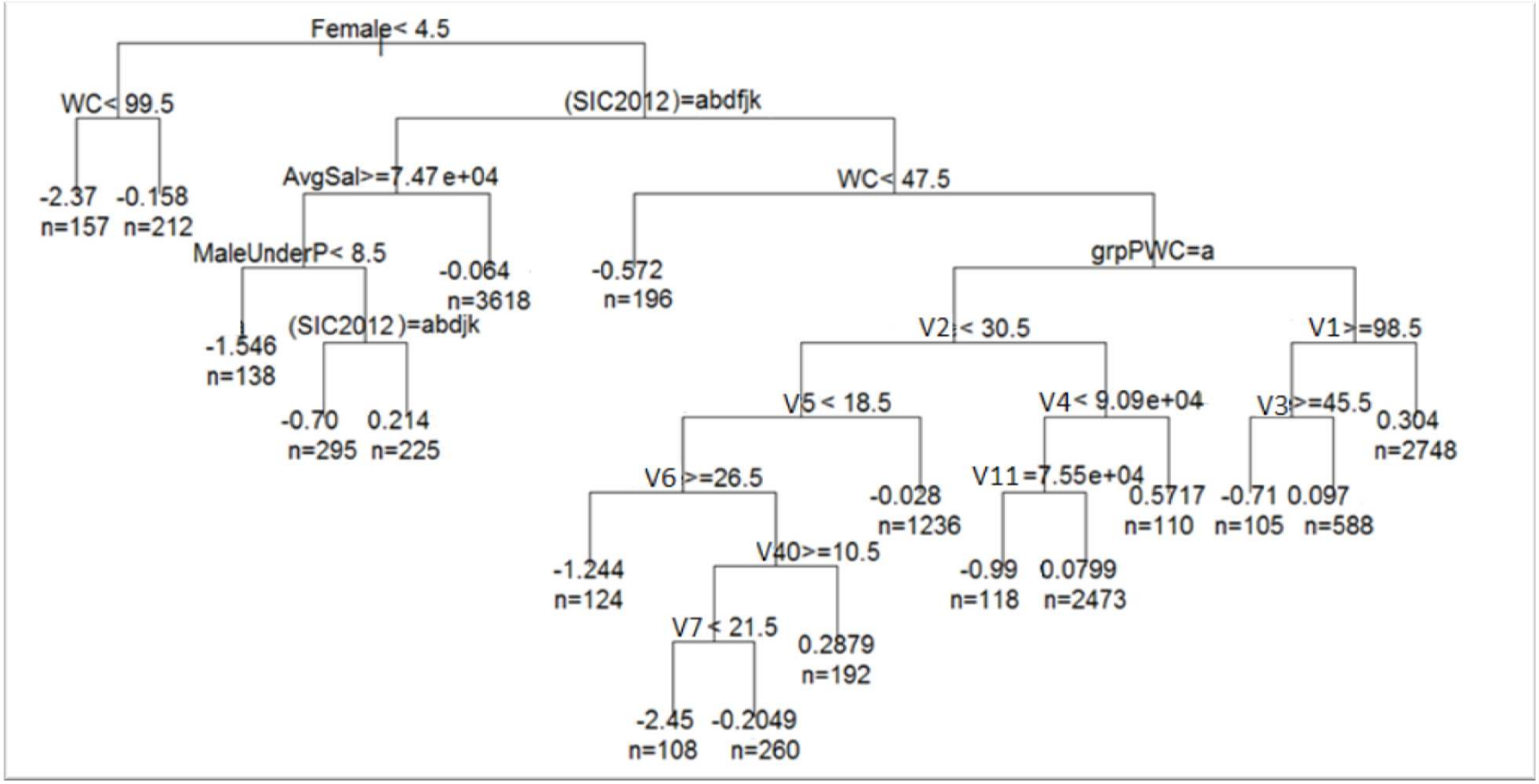
Institute and Faculty of Actuaries

Case Study 2: Pricing Override Model

- Business: US group Long-Term Disability(LTD)
  - About 13k policies, with lives per policies from 10 to 30k
  - Current pricing variables: about 30-40
  - Experience data of past 5 years with >80 variables
  - Major pricing variables: age, gender, industry, location, benefit structure
- Objective
  - To determine additional pricing variables and possible interaction terms (for pricing)
  - To identify groups with experience deviating from pricing assumptions (for UW)

Institute and Faculty of Actuaries

Indicative* CART Model Results



- Easy to develop, interpret and understand; business insights
- Not efficient for linear function; sensitive to noise; over-fitting

Institute and Faculty of Actuaries

* Not actual model results, actual results are client confidential

CART Model Results (2)

- Results improve profit margin and pricing accuracy
  - Useful tool for both pricing and UW of group LTD business
- Model implementation
  - Approved by management team
  - Implemented in Q1'13

Quartile	# of cases	Actual EPM	Model Predicted EPM
1	3230	(28.0%)	(32.0%)
2	3230	(8.8%)	(6.0%)
3	3230	6.3%	2.0%
4	3230	1.7%	1.4%

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Conclusions

Key Messages :

- Predictive Analytics/Modeling is a data-driven process with a broad array of potential applications in insurance.
- Large volumes of good quality, relevant data, is essential for a good result.
- Applications can assist actuaries in their regular jobs.
- Applications to simplified underwriting are proving popular – BUT there are many more applications!
- There are not “off-the-shelf” end-to-end solutions. PM solutions are customized based on specific data and specific needs.
- No two exercises are the same – flexibility of approach is key.

Institute and Faculty of Actuaries

Questions

Comments

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.

Institute and Faculty of Actuaries

Predictive Analytics Projects Worldwide*

* Which the speaker knows about!

UK:
- Basis Setting (mortality, morbidity and lapses)
- Postcode pricing model
- Enhanced experience analysis
- Predictive underwriting on credit rating agency and bank data
- Broker Quality

Europe:
- Predictive underwriting on bancassurance data

South Africa:
- Enhanced Experience Analysis
- Predictive underwriting on bank and credit card data

India:
- Claims Fraud Prediction

Australia:
- Predictive underwriting / cross sell on bancassurance data

USA:
- Pricing override model for group LT disability
- Lapse basis
- Predictive underwriting on Non-Life data
- Term Tail Lapses
- Mortality prediction on credit rating agency data

Asia:
- Predictive underwriting on bancassurance data
- Finer price segmentation
- Propensity to buy
- Cross sell of insurance on bank data