Predictive Analytics: The Convergence of Data Science and Actuarial Practice

Southeastern Actuaries Conference

Justin Fountain
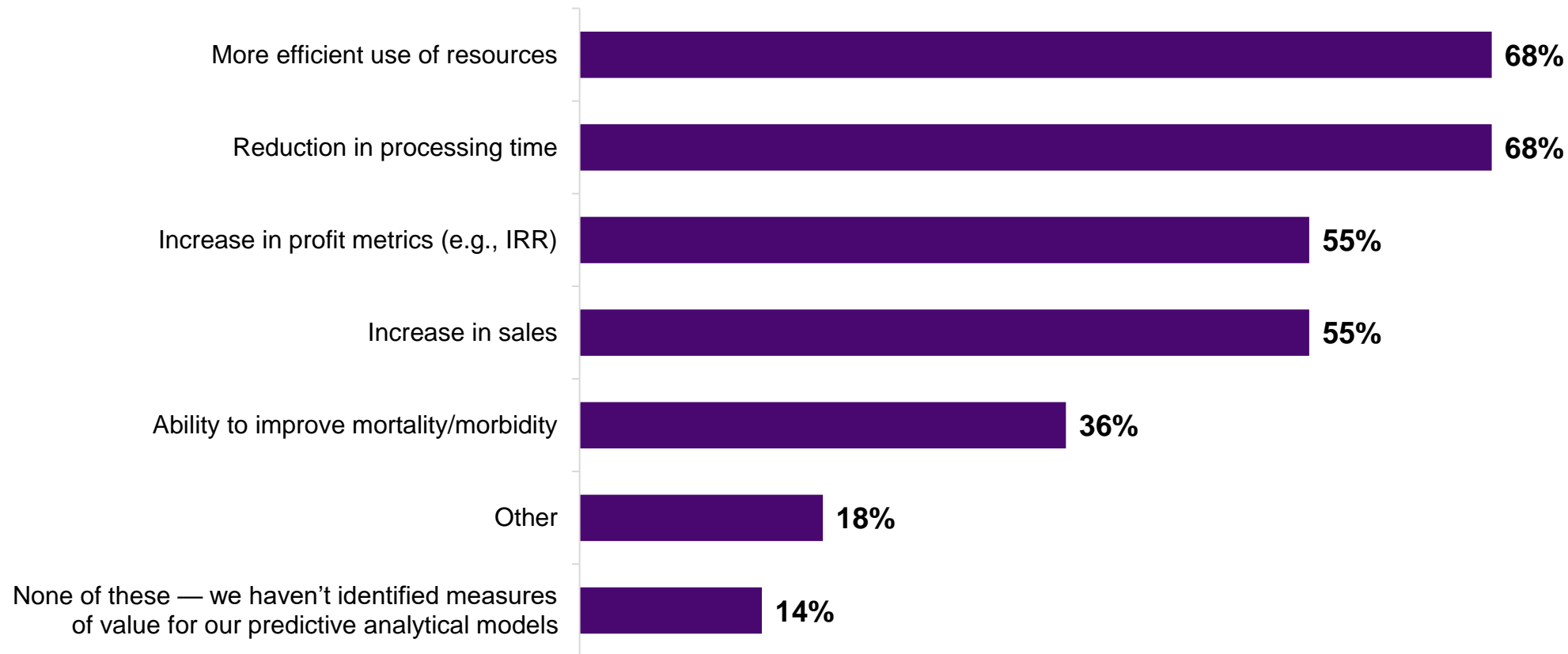
June 20, 2024

wtw

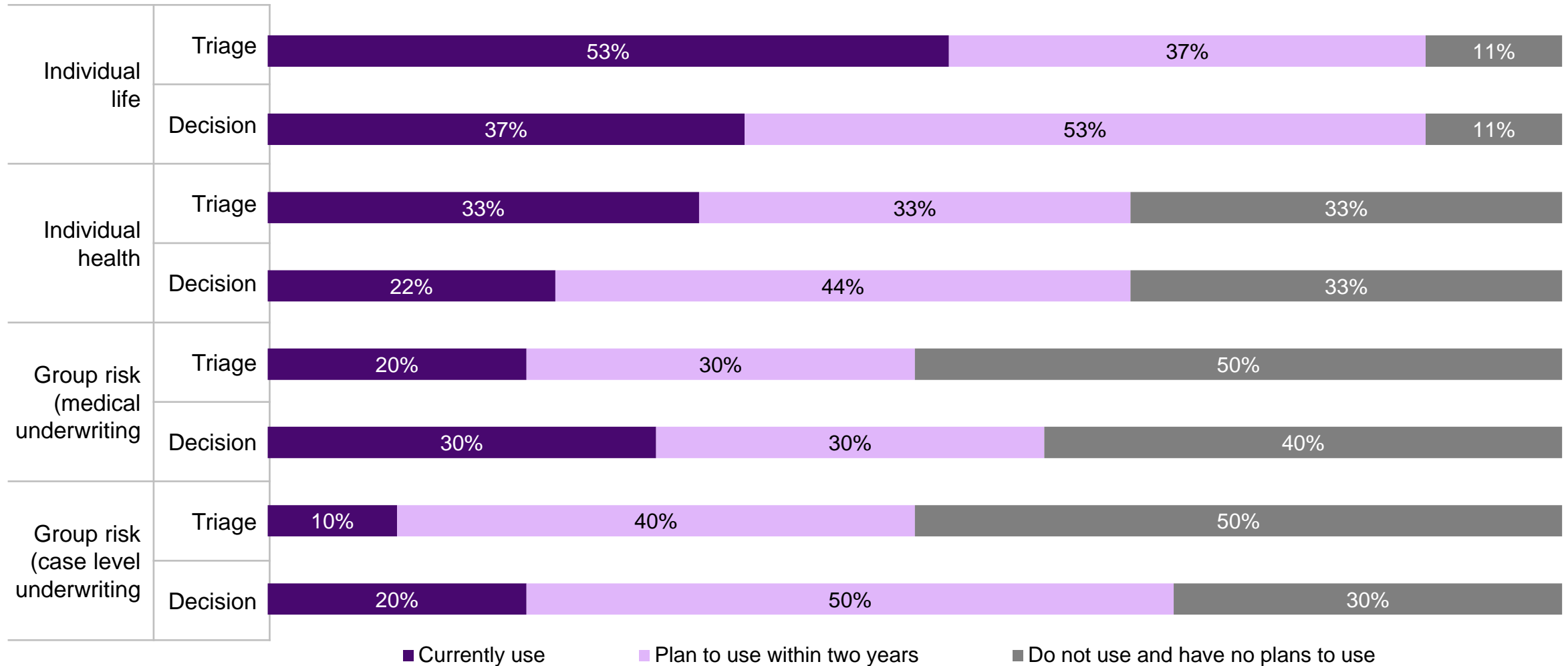# Life insurers realize significant impact from predictive analytics

	Underwriting Triage	Underwriting Decision	Pricing	Mortality / Morbidity	Policyholder Behavior	Lead Generation	Claims Triage / Decision	Inforce Management	Agent Selection
Significant Increase in Sales	50%	57%	50%	33%		62%			
Expansion into New Markets	17%	43%	20%		17%	50%		20%	17%
Increased Placement Rates	60%	71%				43%		33%	14%
Increased Profitability	38%	51%	38%	22%	11%	43%	50%	14%	17%
Reduced acquisition Expenses	66%	88%							20%

80-100	60-79	40-59	20-39	1-19

wtw

Value of predictive analytics initiatives to life insurers

Category	Percentage
More efficient use of resources	68%
Reduction in processing time	68%
Increase in profit metrics (e.g., IRR)	55%
Increase in sales	55%
Ability to improve mortality/morbidity	36%
Other	18%
None of these — we haven't identified measures of value for our predictive analytical models	14%

Predictive analytics is becoming necessary to remain viable

		Currently use	Plan to use within two years	Do not use and have no plans to use
Individual life	Triage	53%	37%	11%
	Decision	37%	53%	11%
Individual health	Triage	33%	33%	33%
	Decision	22%	44%	33%
Group risk (medical underwriting	Triage	20%	30%	50%
	Decision	30%	30%	40%
Group risk (case level underwriting	Triage	10%	40%	50%
	Decision	20%	50%	30%

■ Currently use   ■ Plan to use within two years   ■ Do not use and have no plans to use

wtw

Where do companies believe predictive analytic skillsets are found?

Education and training
- Neither 4%
- Actuary 8%
- Both 56%
- Data scientist 32%
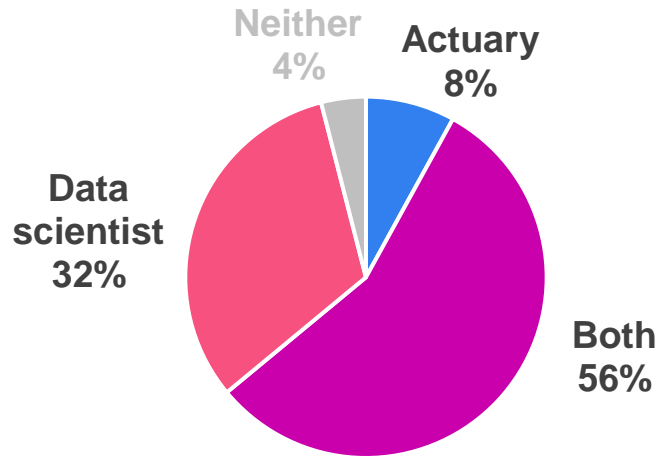
Insurance knowledge
- Both 8%
- Neither 4%
- Actuary 88%
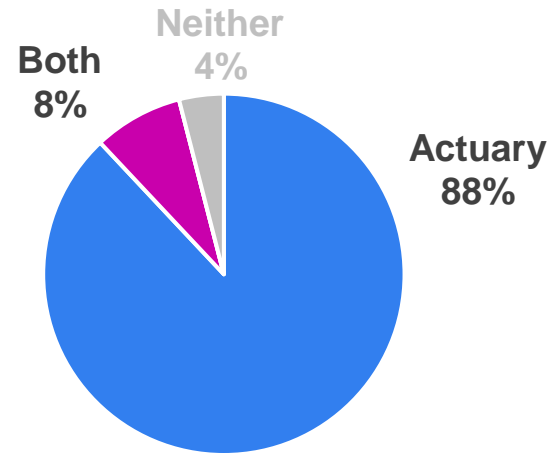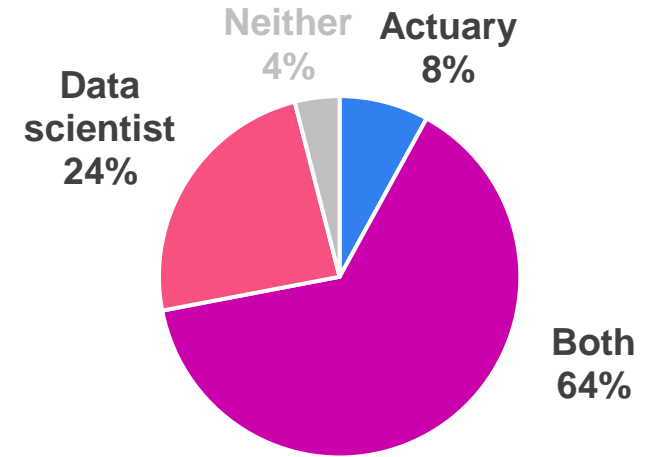
Mathematical and statistical aptitude
- Neither 4%
- Actuary 8%
- Data scientist 24%
- Both 64%
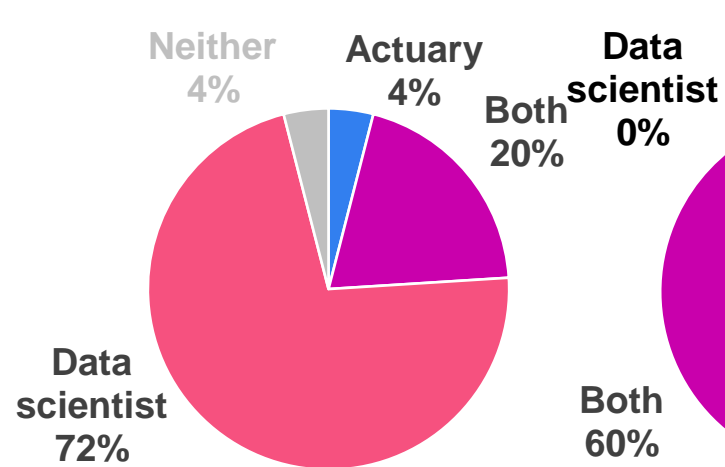
Technical and software skills
- Neither 4%
- Actuary 4%
- Both 20%
- Data scientist 72%
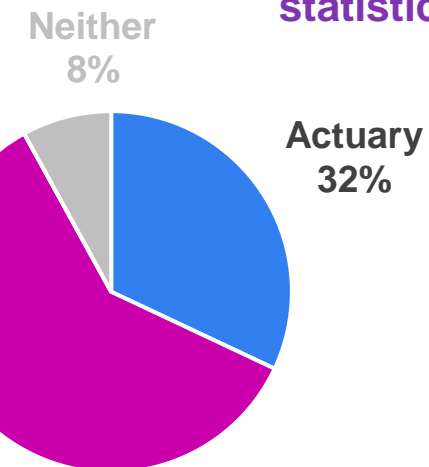
Communication skills
- Data scientist 0%
- Neither 8%
- Actuary 32%
- Both 60%

Generative AI

85%
Insurance execs plan to increase investment in generative AI within 2024[1]

9%
Insurance execs confident they have technology in place to leverage generative AI[1]

15%
Increased revenue for insurers[2]

10%
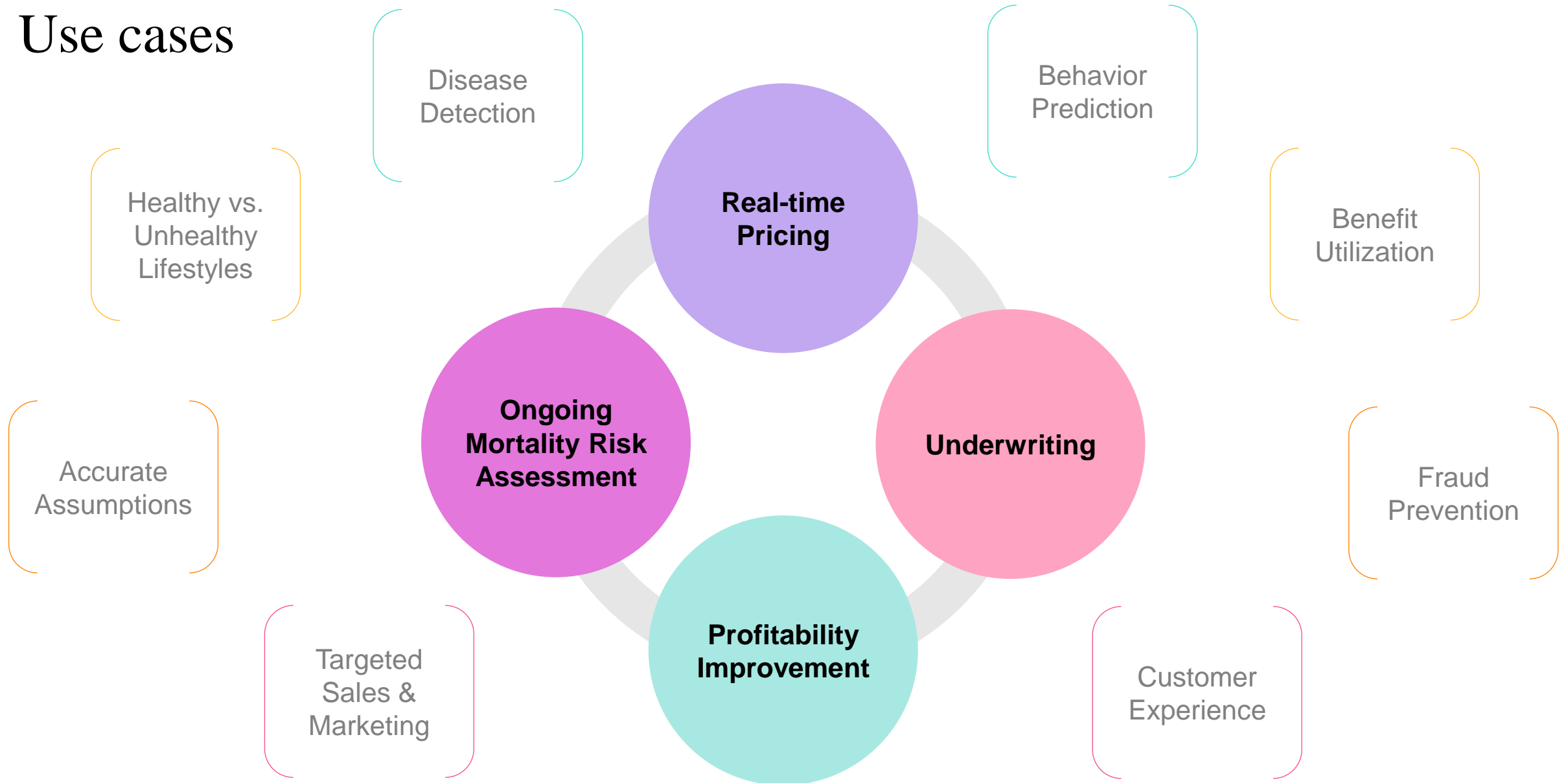Reduction in expenses for insurers[2]

Sources:
1 https://www.accenture.com/us-en/insights/pulse-of-change
2 https://www.bain.com/insights/its-for-real-generative-ai-takes-hold-in-insurance-distribution/

Use cases

Disease Detection

Behavior Prediction

Healthy vs. Unhealthy Lifestyles

Benefit Utilization

Real-time Pricing

Ongoing Mortality Risk Assessment

Underwriting

Accurate Assumptions

Fraud Prevention

Profitability Improvement
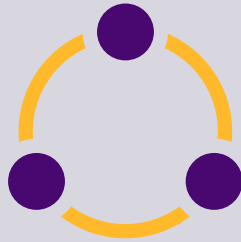
Targeted Sales & Marketing

Customer Experience

wtw

Business objectives and strategic goals for pricing

Increase in margin and revenue

Consistency of decisions

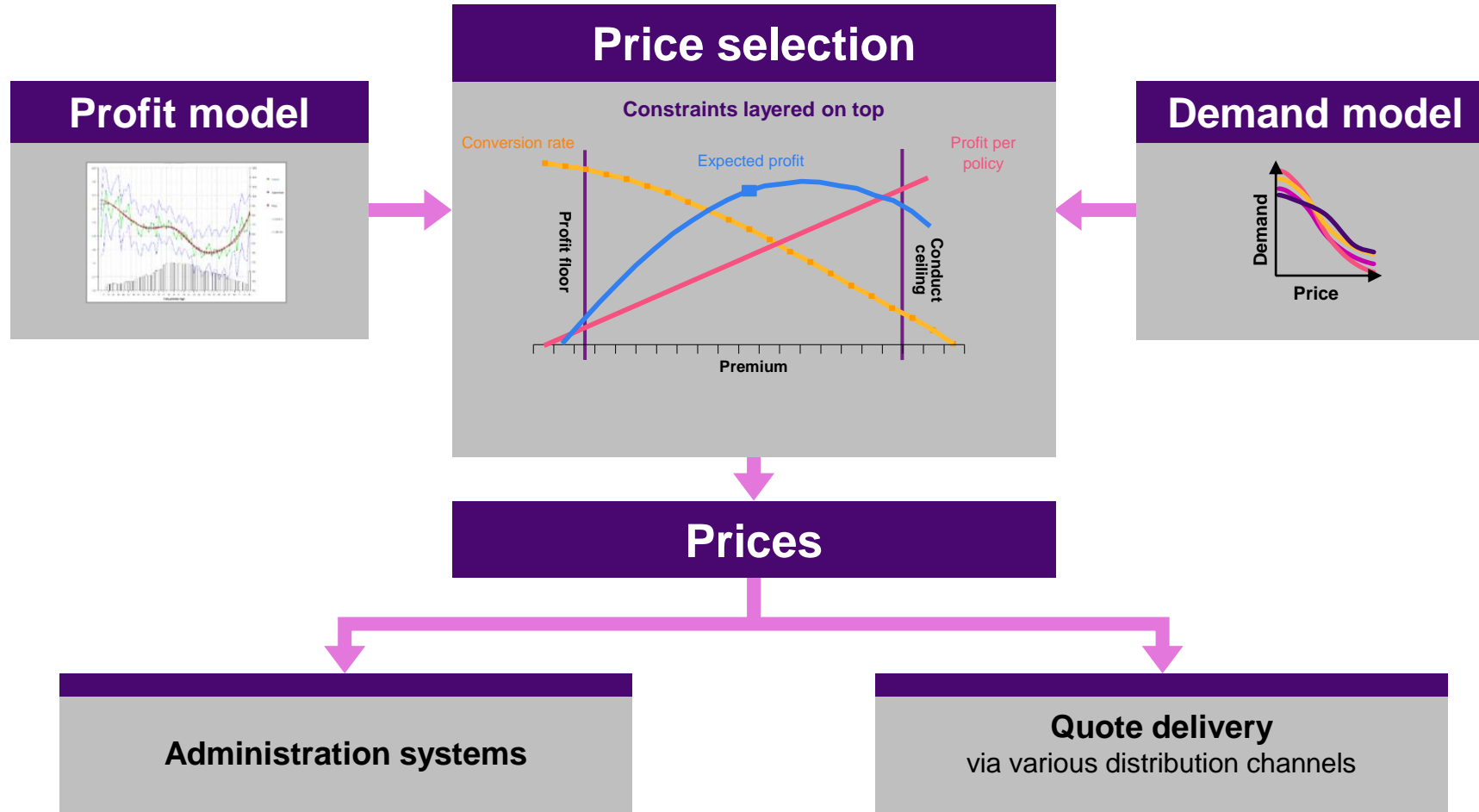**Reduce risk** of dependency on competitor prices

Be more strategic by increasing understanding and control of your business

Emerging life experience demonstrates an increase in value of **2% to 5%** of annual premiums from initial dynamic pricing implementations, with the potential for up to **10%**
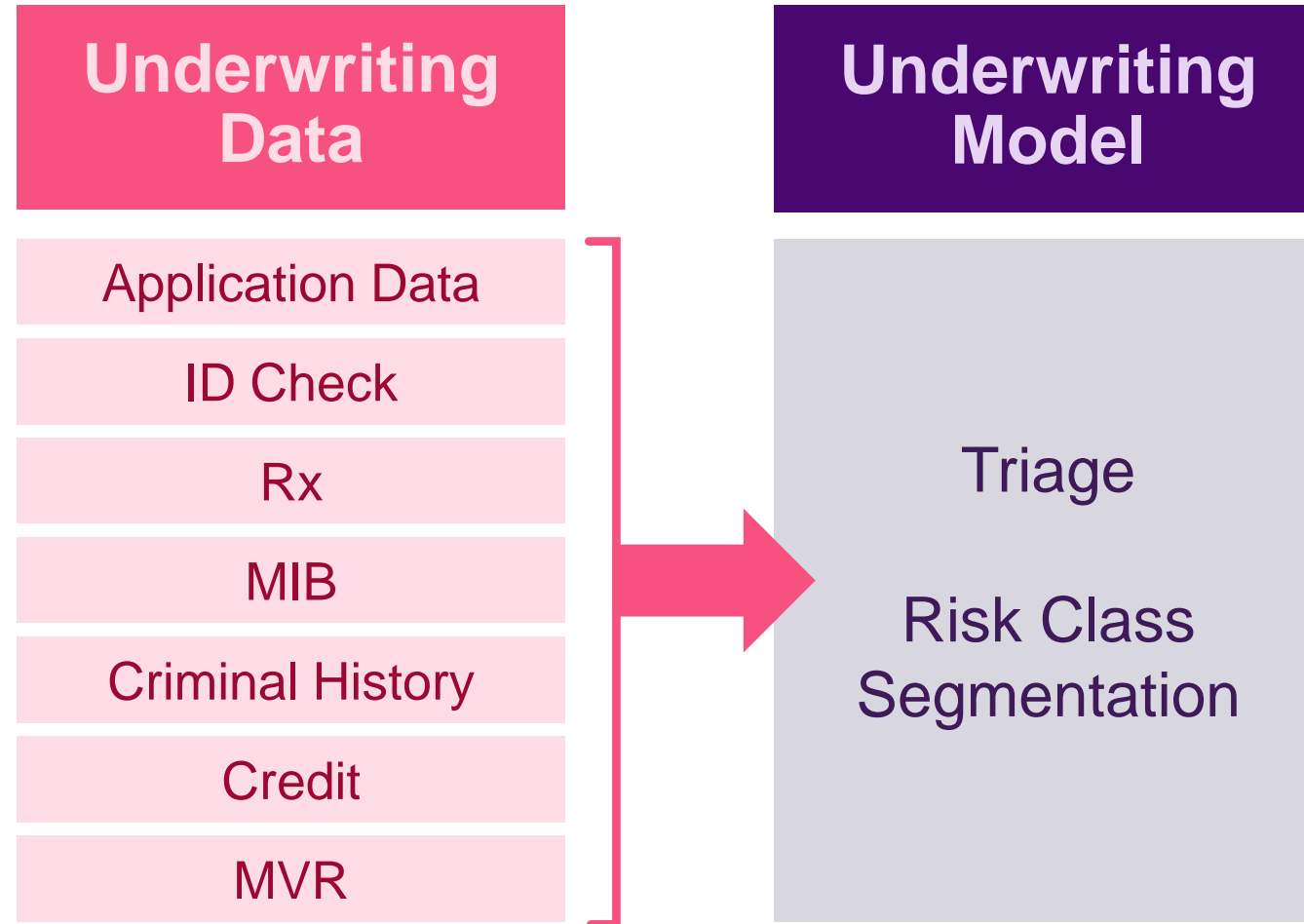
Dynamic pricing

Business objectives and strategic goals for underwriting

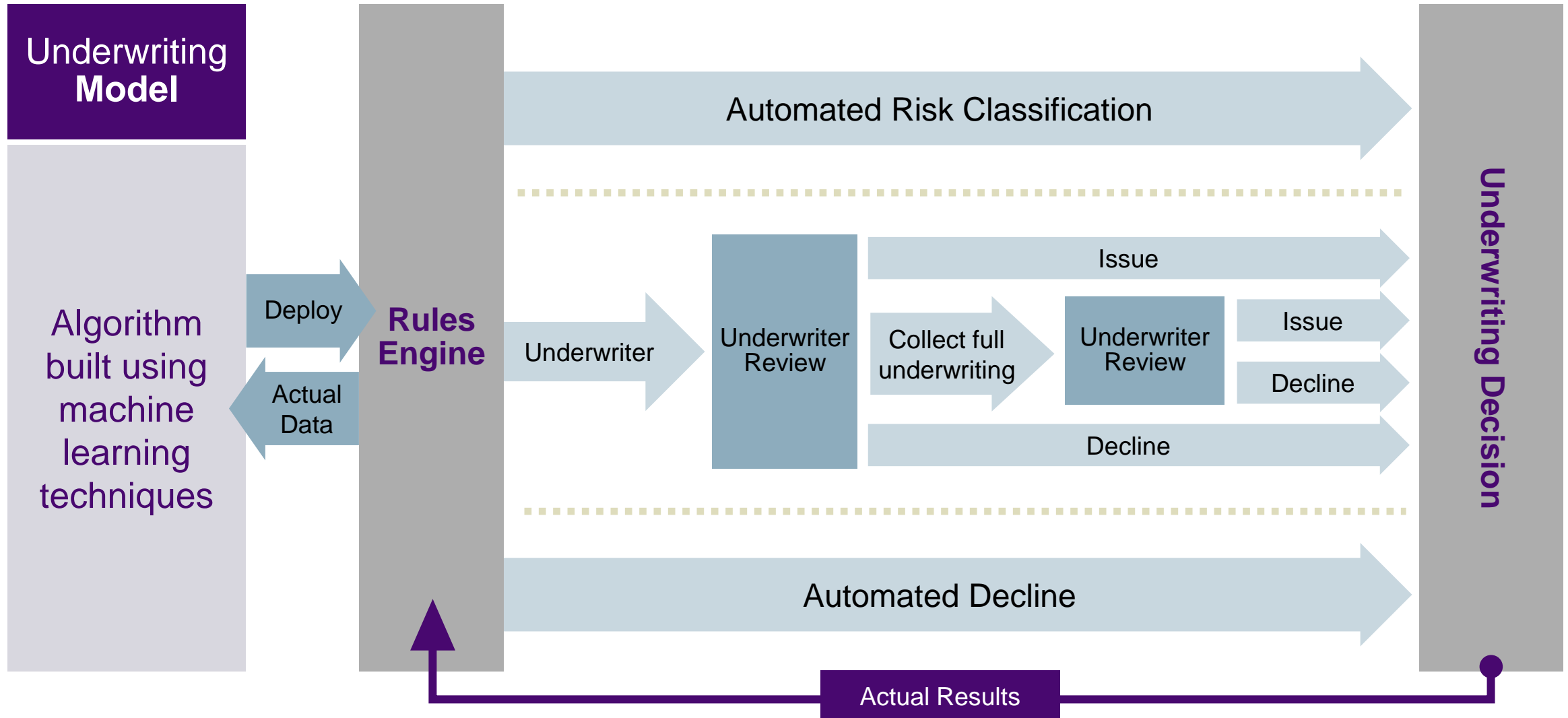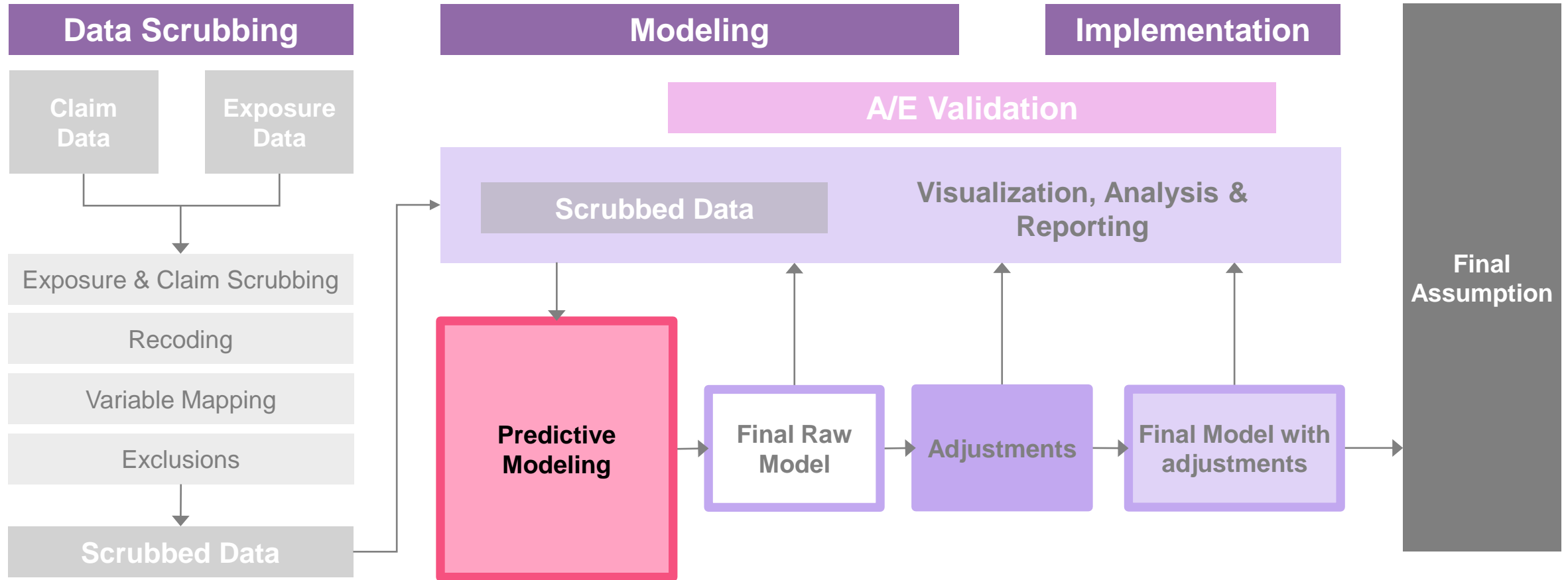What are you trying to achieve by revolutionizing your underwriting?

Increase Profitability by Selecting the Best Risks

Improve the Customer Experience

Reduce Underwriting Expenses

Remain Competitive

wtw

Modern (accelerated) underwriting

Software can be used to analyze data

Underwriting Data
Application Data
ID Check
Rx
MIB
Criminal History
Credit
MVR

Underwriting Model

Triage

Risk Class Segmentation

Modern (accelerated) underwriting

Experience analysis

Data Scrubbing	Modeling	Implementation

Data Scrubbing

Claim Data	Exposure Data

Exposure & Claim Scrubbing

Recoding

Variable Mapping

Exclusions

Scrubbed Data

Modeling

A/E Validation

Scrubbed Data

Visualization, Analysis & Reporting

Predictive Modeling → **Final Raw Model** → **Adjustments** → **Final Model with adjustments**

Implementation

Final Assumption

Real world example of predictive analytics benefits

Classical vs. Predictive Analytics
TOAMS 5 Industry Mortality Study
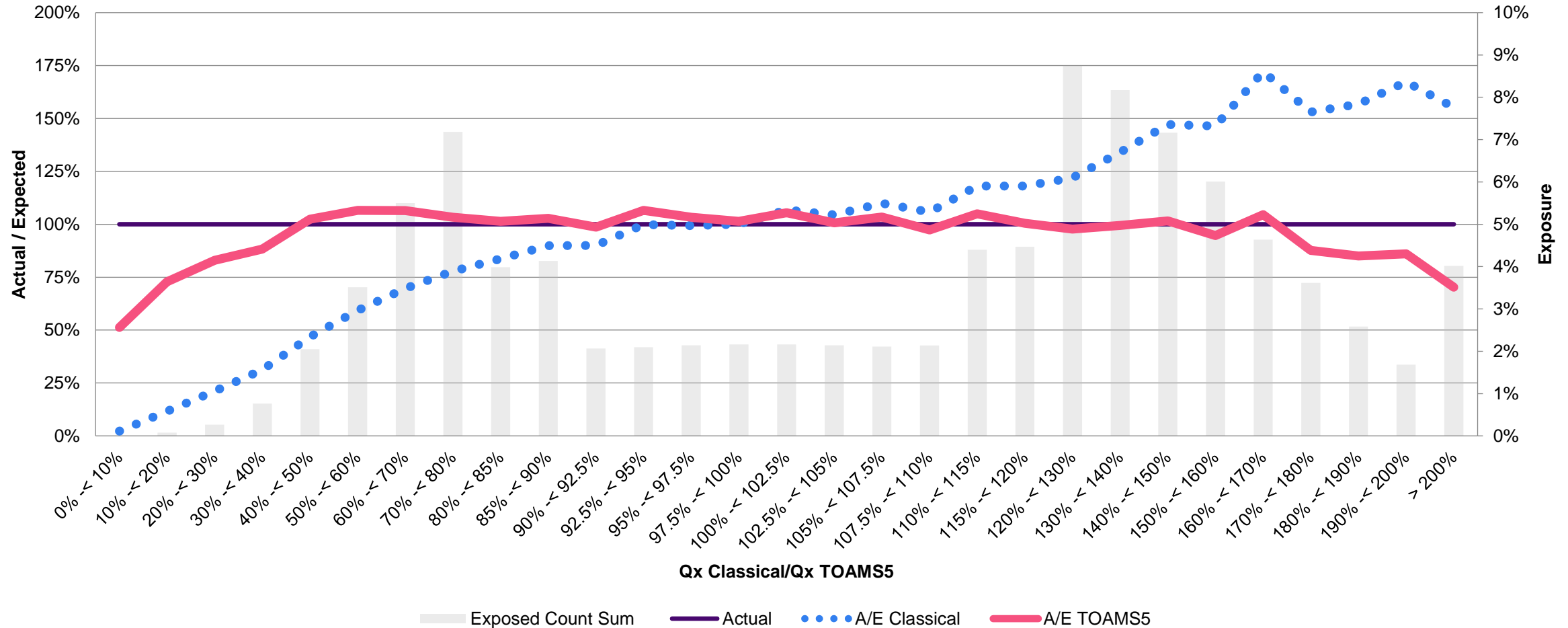
Desired characteristics of advanced analytics tools

What are you looking for?

Flexibility

Fewer human hours needed

Strong governance

Ease of use

Predictive Analytics

What happened?

Descriptive Analytics

What might happen next?

Predictive Analytics

What would happen if I do this?

Prescriptive Analytics

Predictive analytics is the use of various modeling techniques (including machine learning) to analyze historical data to identify patterns that might predict future outcomes
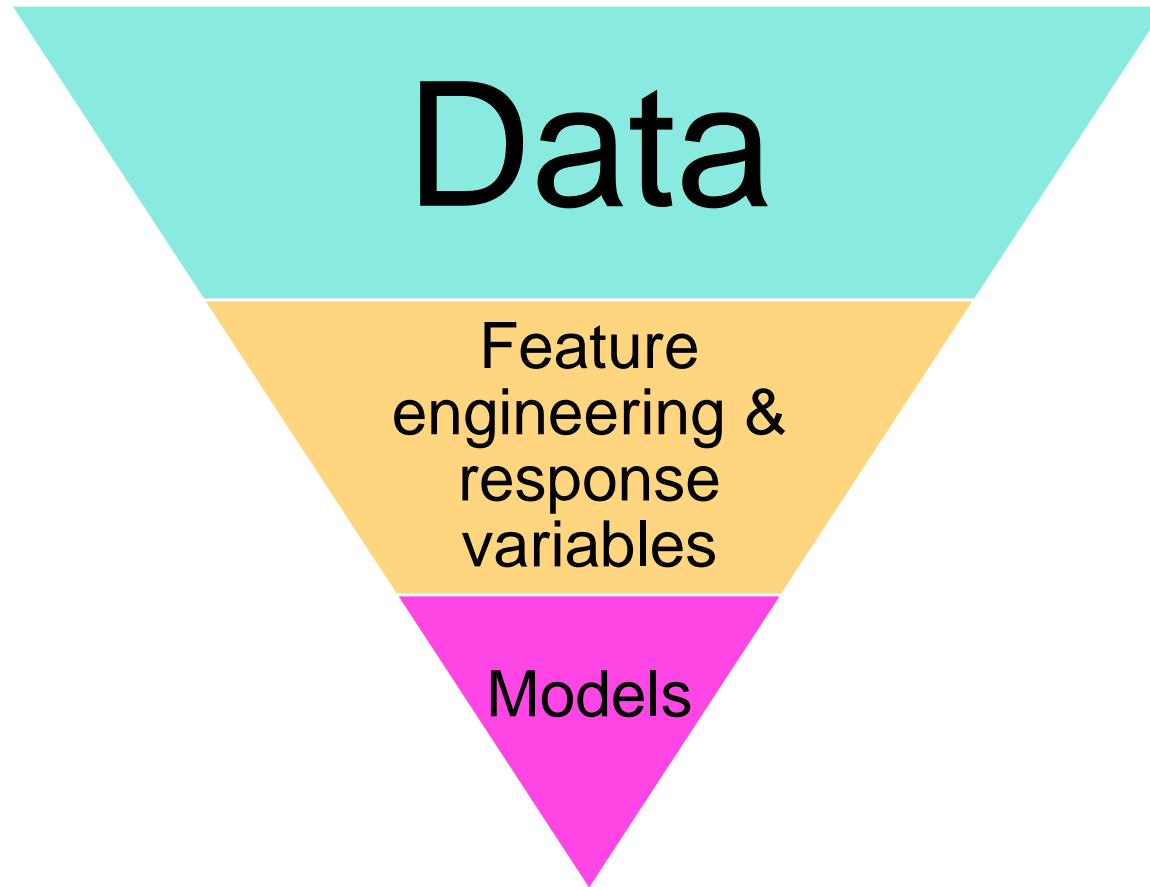
Predictive analytics cannot solve all problems

It's important to remember that predictive analytics only works if the problem is ___**actually solvable**___ with the data that you have
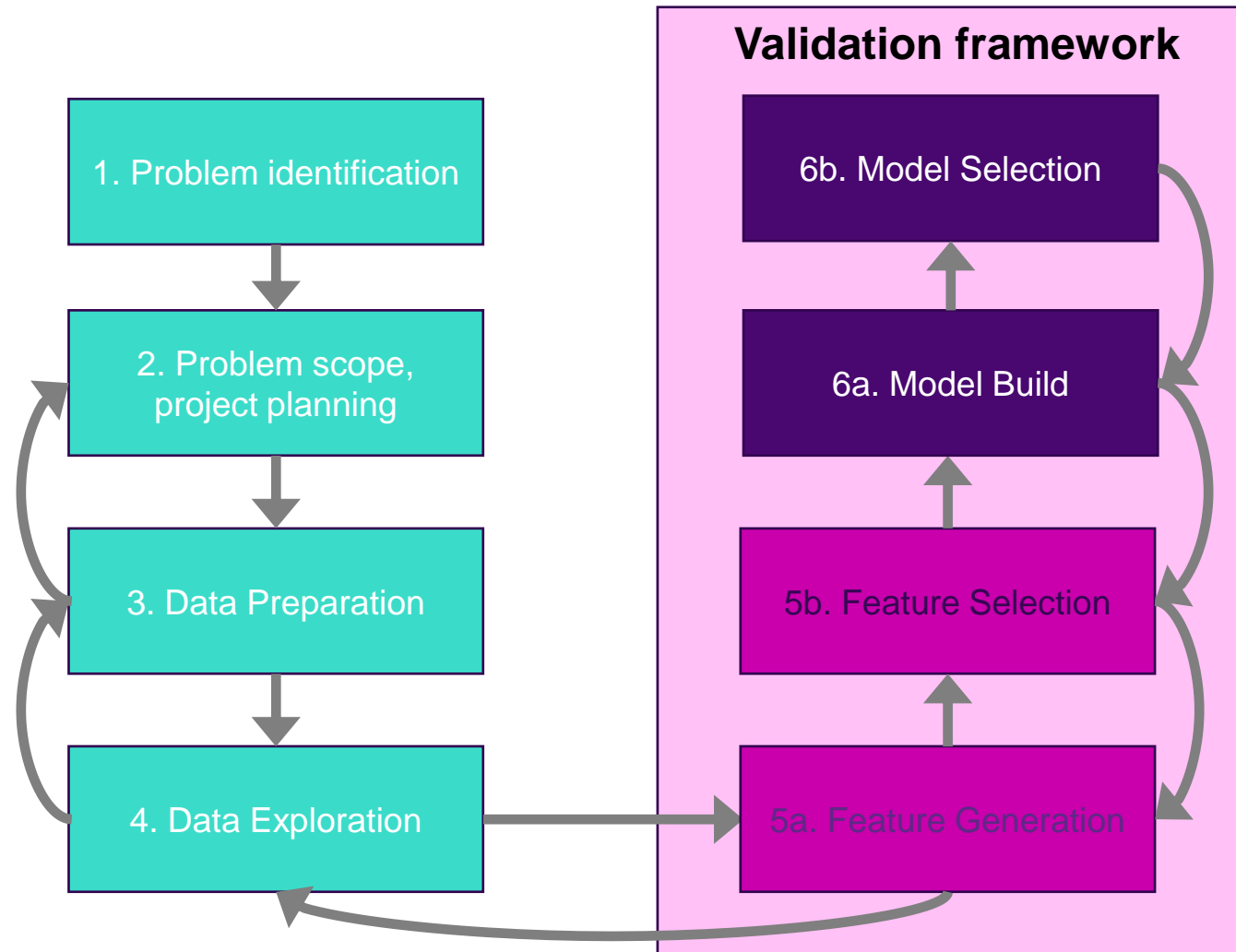
Key Considerations:

- You can clearly identify and define a business issue that needs to be addressed
- You can address the issue with a few well-defined questions
- You have plenty of high-quality data that can be used to answer these questions
- You are certain the predictions will drive actions
- You are confident that it is better than any existing process or approach
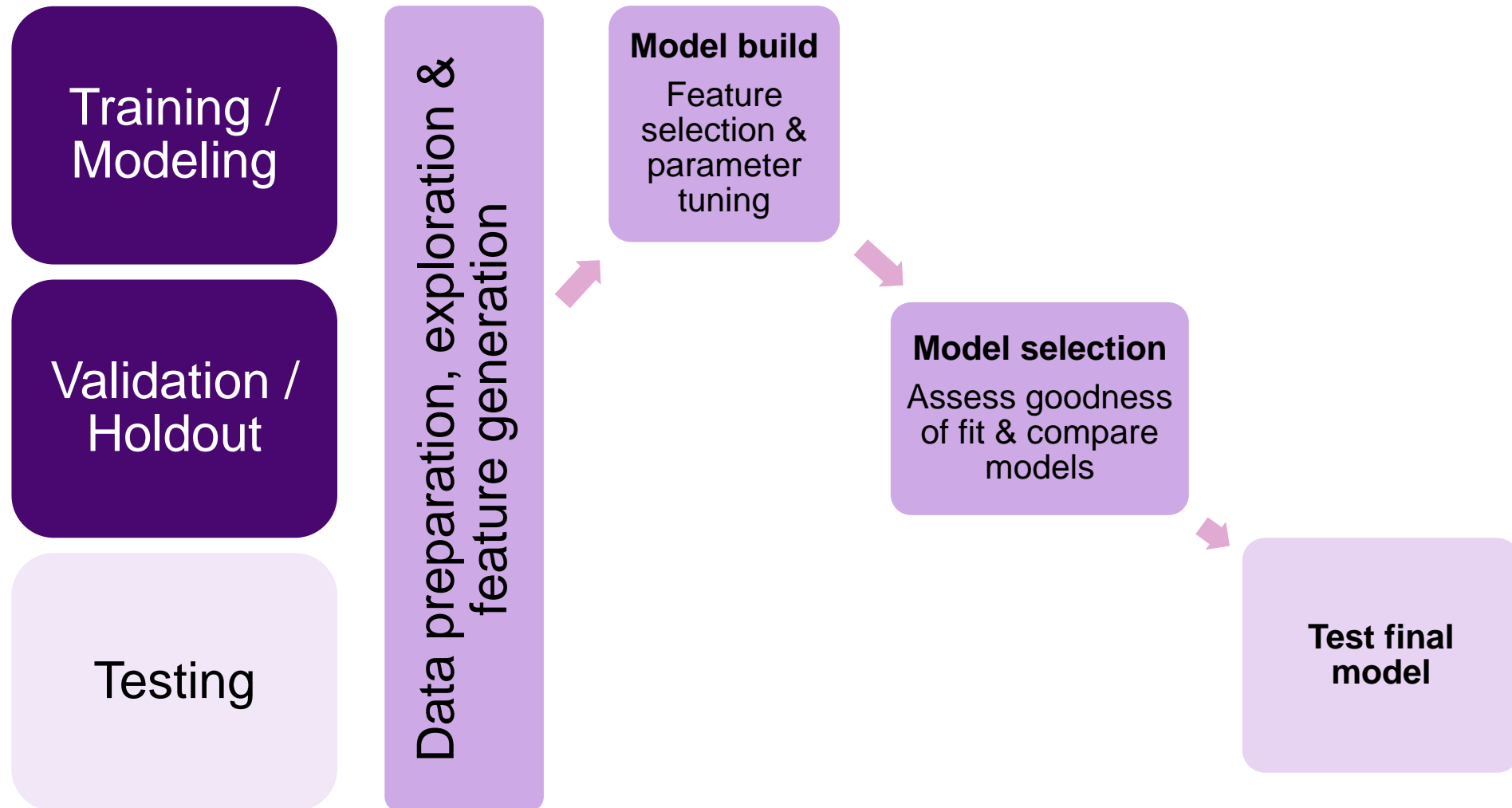- You can continue to monitor and update the models when new data is available
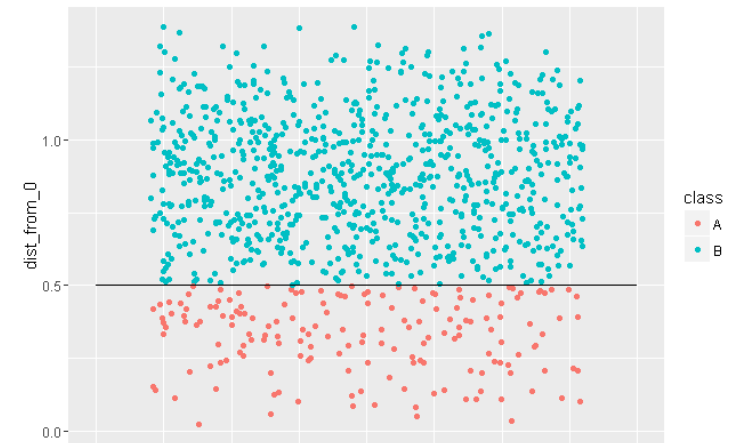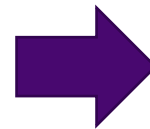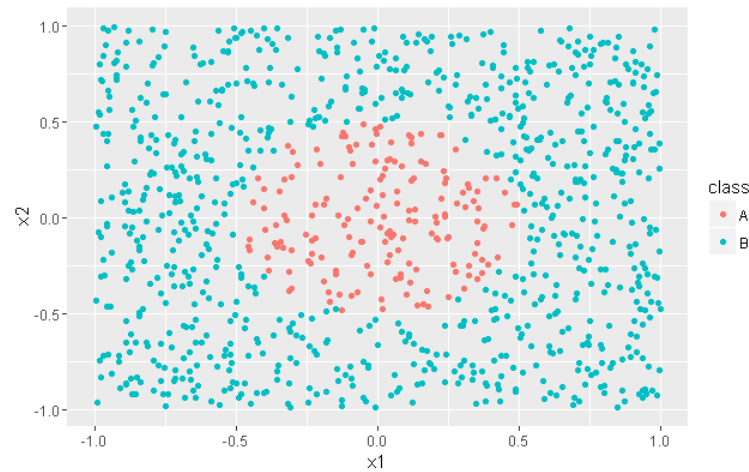
Predictive analysis process

Iterative process

Validation framework

Training / Modeling

Validation / Holdout

Testing

Data preparation, exploration & feature generation

Model build
Feature selection & parameter tuning

Model selection
Assess goodness of fit & compare models

Test final model

Feature development
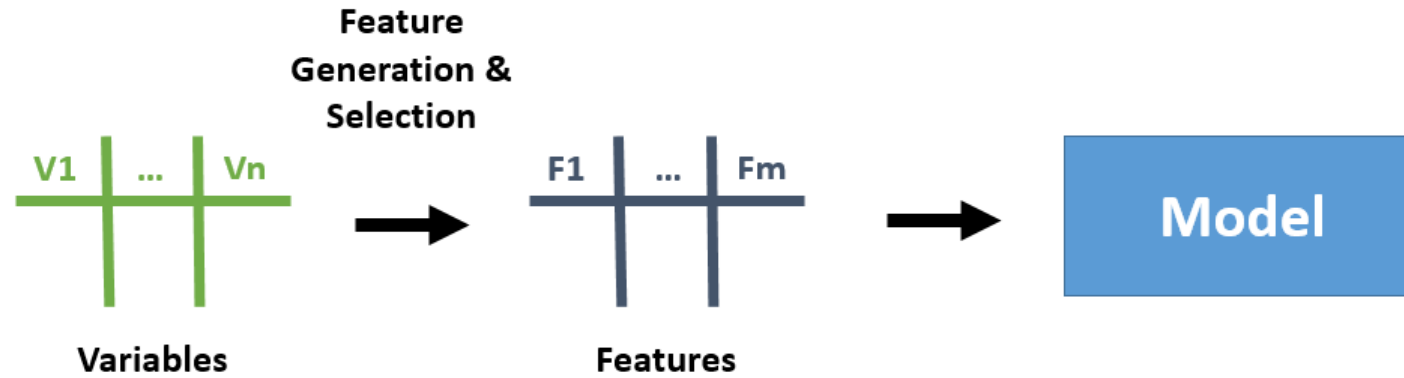
Some key terms to remember

Weight

Exposure

Response
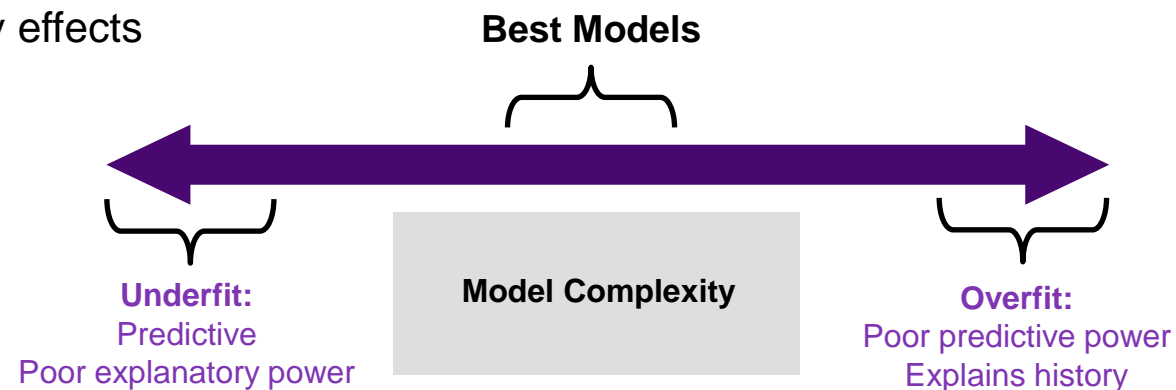Target variable
Outcome
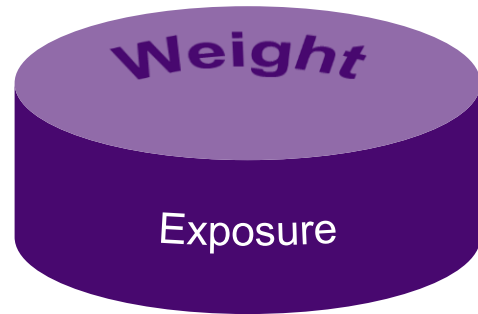
What is the goal of model selection?

The goal is to produce an analysis that explains recent historical experience and is predictive of future experience

1. Separate the signal from the noise   **Data   =   Signal   +   Noise**

2. Balance predictive and explanatory effects

Best Models

Underfit:
Predictive
Poor explanatory power

Model Complexity

Overfit:
Poor predictive power
Explains history

The risk in simplicity is to miss signal whereas the risk in complexity is to overreact to the noise
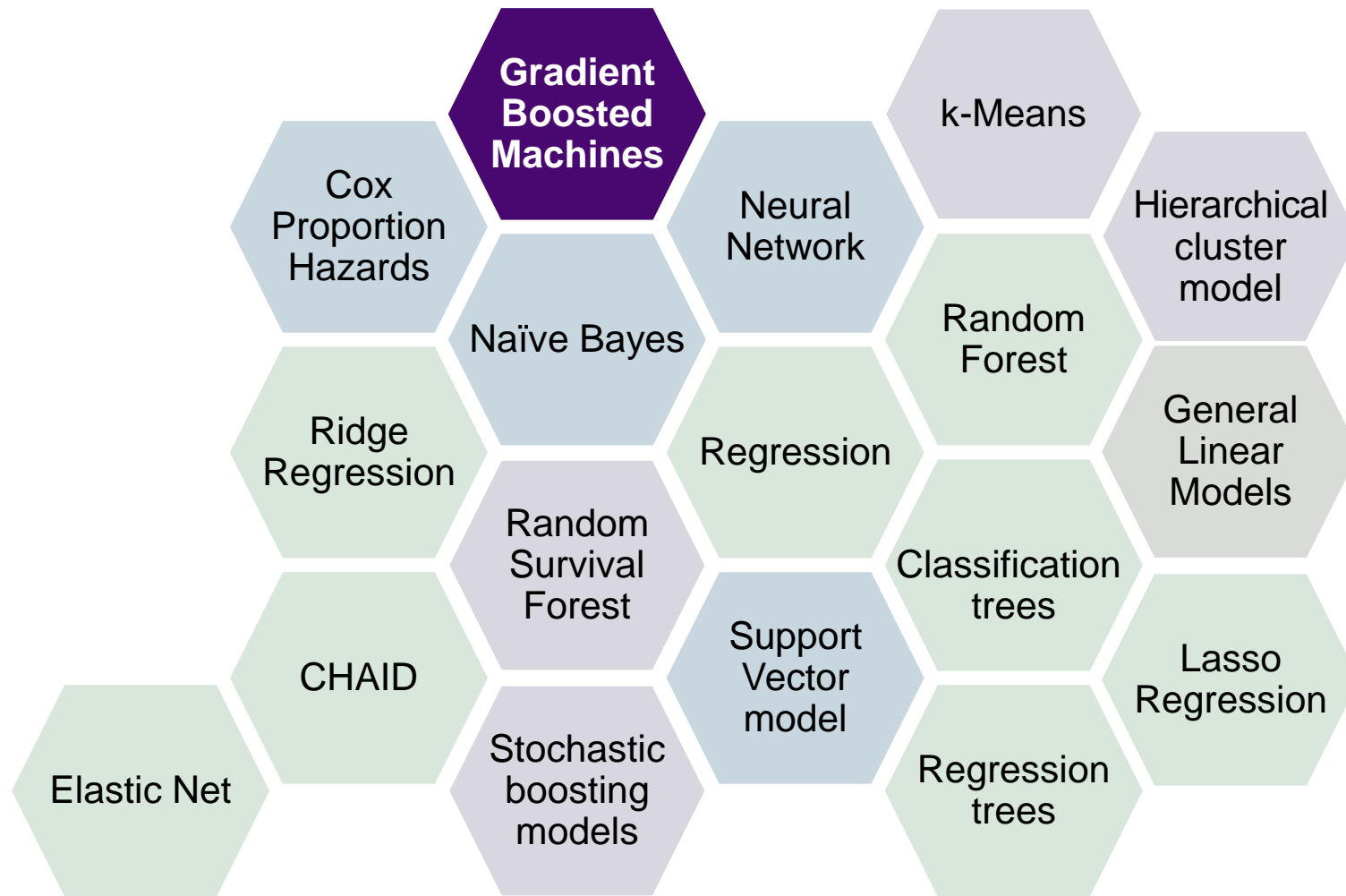
What makes a good model?

Predictive Power

All of these factors go into model selection

Remember…

"…all models are wrong, but some models are useful."

- George Box

wtwco.com

wtw

Machine learning

Generalized linear models (GLMs)

GLMs take the following form:

$$y \quad = \quad h(\text{Combination of variables}) \quad + \quad \text{Error}$$

Link function ($g=h^{-1}$) determines how variables are related

- Include variables that are predictive; exclude those that are not
- Simplify, if necessary, using groups and curves
- Include combinations of variables if necessary
- Best combination is determined by experiment, using a combination of statistics and judgment

Reflects the variability of the underlying process and can be any distribution within a broad family (exponential distributions)

Process	Link Function	Error Structure
Frequency	Log	Poisson
Severity	Log	Gamma
Yes/No Behavior	Logit	Binomial

Decision trees

- Decision Trees (Classification and Regression Trees) determine a set of rules that segment observations
- The predicted value is the average value within the terminal node
- The model stops when no further splitting will improve results, or some stopping condition is satisfied

Survival of passengers of the Titanic
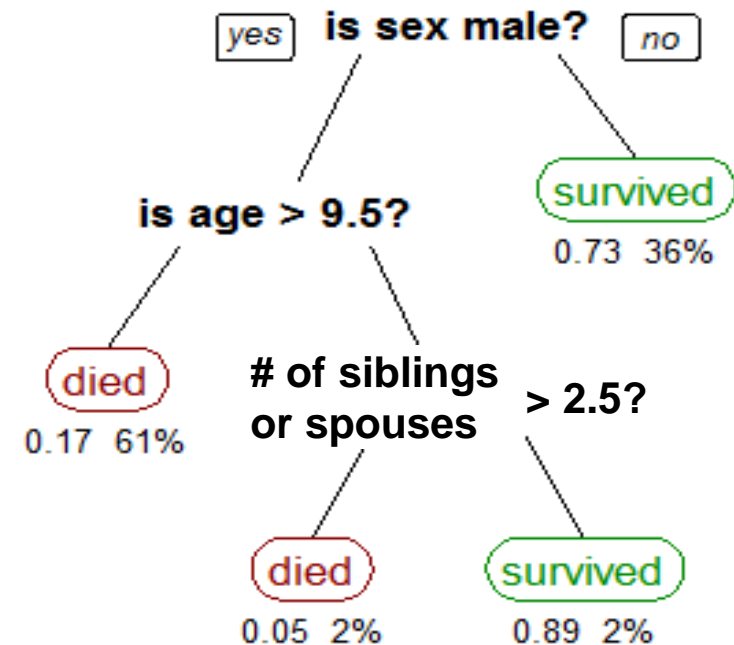
Random forests

Roughly speaking, fitting a random forest involves:

- Obtaining numerous random samples of the data (with replacement)
- Fitting a simple tree on each sample
- The model is an average of these trees

The idea is that the combination of simple trees fit on different samples avoids overfitting to the data, and is more predictive than any single tree
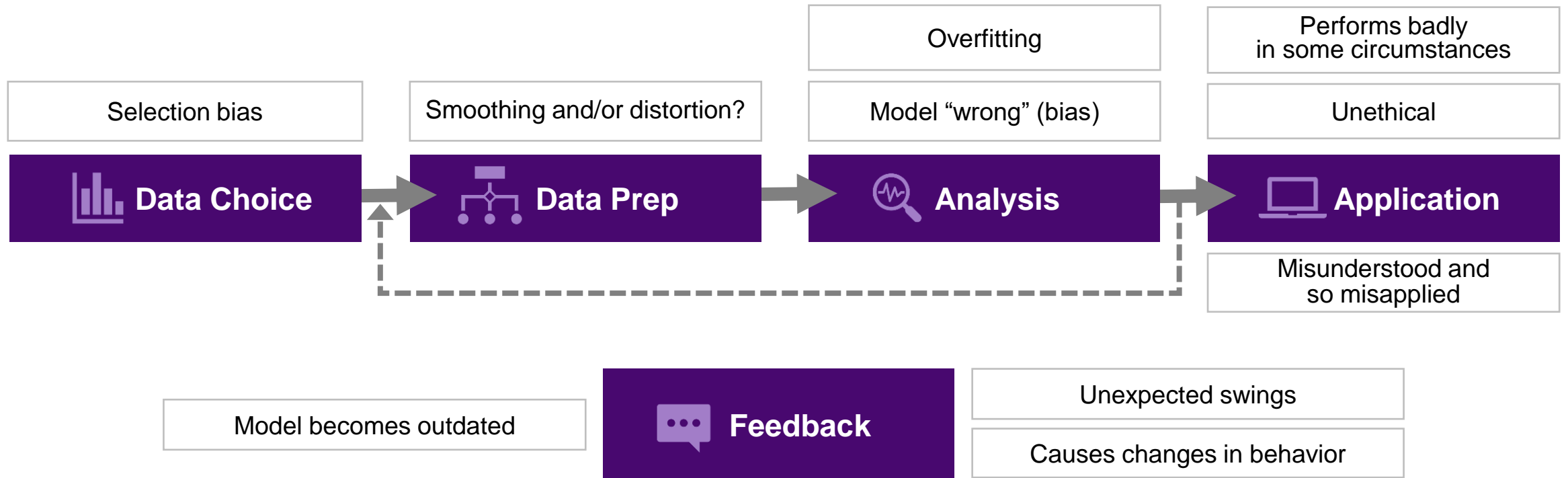
Model comparison

	GLM	Penalized Regression	MARS / Earth	GBMs	Trees	Neural Networks	Random Forests	Support Vector Machine
Predictive power	4	3	3	5	1	4	4	4
Interpretation	5	3	5	1	5	1	1	1
Table implementation	5	5	5	1	5	3	3	4
Stability	5	5	3	3	1	3	3	3
Execution speed	5	5	5	1	5	5	5	3
Analytical time & effort	2	4	5	3	5	1	1	4

5 = Most suitable → *1 = Least suitable*

What could go wrong?

The human element is still critical

		Overfitting	Performs badly in some circumstances
Selection bias	Smoothing and/or distortion?	Model "wrong" (bias)	Unethical

Data Choice → **Data Prep** → **Analysis** → **Application**

Misunderstood and so misapplied

Model becomes outdated   **Feedback**   Unexpected swings

Causes changes in behavior

wtwco.com

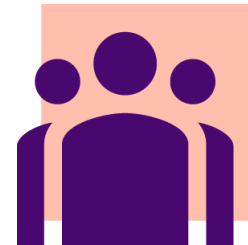# What are the challenges to consider?

Expertise

Computer/IT Systems
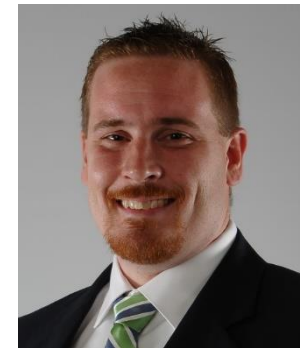
Privacy

Acceptance

More information

Links

soa.org/sections/pred-analytics-futurism/

soa.org/research/topics/pred-analytics-topic-landing/

soa.org/research/topics/life-exp-study-list/

Actuarial Standards of Practice

- Setting Assumptions (proposed)
- ASOP 2 – Nonguaranteed Elements
- ASOP 54 – Pricing
- ASOP 23 – Data Quality
- ASOP 41 – Actuarial Communications

Justin Fountain
WTW
5 Concourse Pkwy
Atlanta, GA 30306
Direct: +1 678 684 0559
Mobile: +1 770 826 9339

justin.fountain@wtwco.com