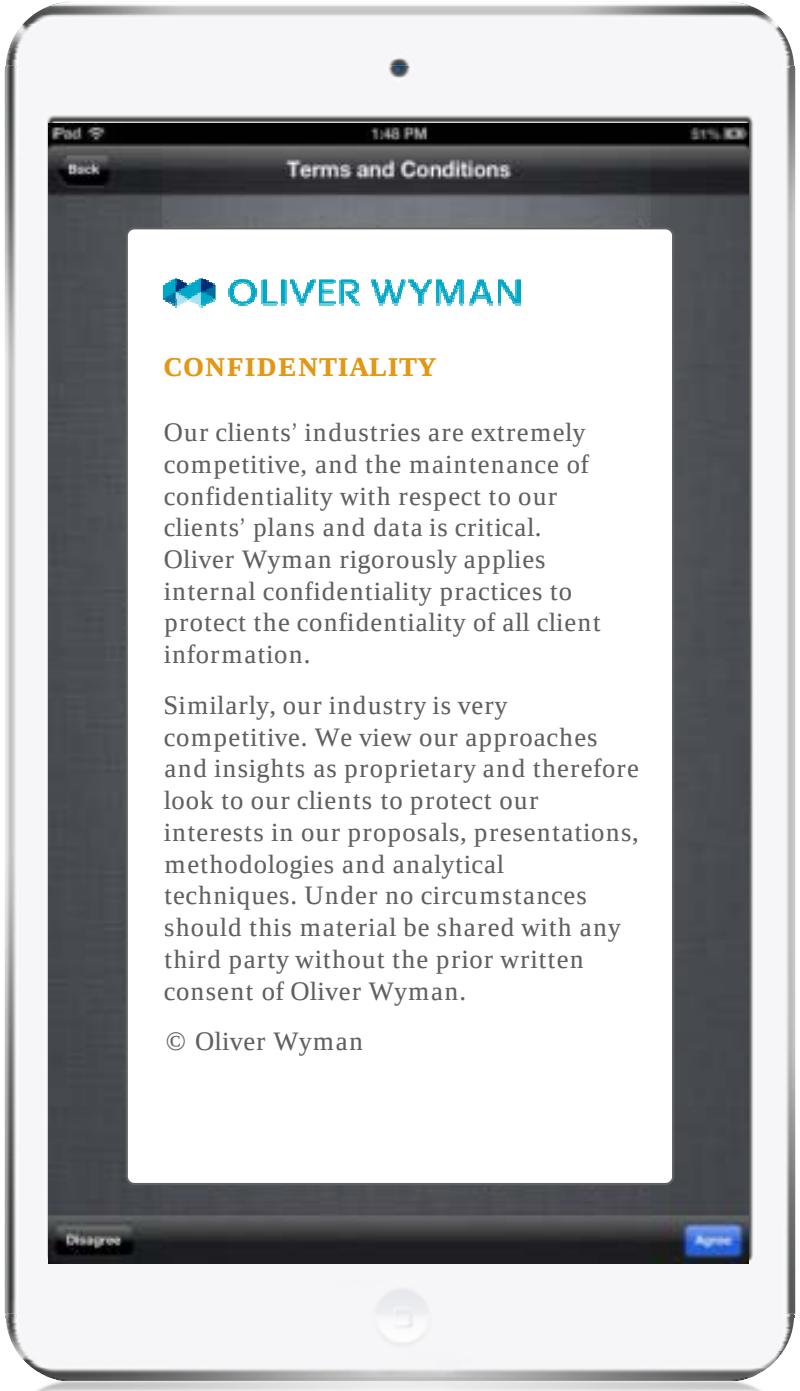
OLIVER WYMAN

ADVANCED ANALYTICS FOR RISK MANAGEMENT

AUGUST 8, 2018

DIGITAL

MARSH & McLENNAN COMPANIES


Pad
1:48 PM
51%
Back
Terms and Conditions
OLIVER WYMAN
CONFIDENTIALITY
Our clients' industries are extremely competitive, and the maintenance of confidentiality with respect to our clients' plans and data is critical. Oliver Wyman rigorously applies internal confidentiality practices to protect the confidentiality of all client information.
Similarly, our industry is very competitive. We view our approaches and insights as proprietary and therefore look to our clients to protect our interests in our proposals, presentations, methodologies and analytical techniques. Under no circumstances should this material be shared with any third party without the prior written consent of Oliver Wyman.
© Oliver Wyman
Disagree
Agree

Presenter

Mikko Lehtonen
Engagement Manager
Finance and Risk, Singapore

mikko.lehtonen@oliverwyman.com

- **Explores banking opportunities**
- **Establishes banking alliances**
- **Advocates banking policies**
- **Trains banking leaders**

Oliver Wyman Group is a part of Marsh & McLennan Companies

- **Revenue 2016: US$ 13.2 BN**
- **Staff: 60,000**
- **Clients in more than 130 countries**
- **Listed on the New York Stock Exchange**

Marsh	Guy Carpenter	Mercer	Oliver Wyman Group
Insurance broker and risk advisor • Risk management, consulting and transfer • Financial solutions and insurance program management services	**Risk and reinsurance specialist** • Reinsurance and risk management • Contract management, claims handling and fiduciary accounting	**Human resource consulting** • Personnel strategy • HR systems • Compensation • Incentive systems • Communication	**Management and economic consulting** • **Strategy** • **Organization** • **Operations** • **Risk management** • **Transformation management**
• Revenue 2016 US$5.98 BN • Staff: 30,000+ • Clients in over 130 countries	• Revenue 2016 US$1.14 BN • Staff: 2,300+ • Offices in over 50 countries	• Revenue 2016 • US$4.32 BN • Staff: 21,000+ • Offices and operations in more than 40 countries	• **Revenue 2016 US$1.79 BN** • **Staff: 4,500+** • **50+ Offices in ~ 30 countries**

Agenda for today

- Recap from previous webinar
- Machine Learning in credit modelling
- Live demo
- Implications
- QnA

1 Recap from previous webinar

Technology and data are currently driving change in risk management across the banking sector

Technology-related enablers for change in risk management

I. MACHINE LEARNING

A. Data science and advanced analytics (i.e. machine learning) provide new and sharper insights

B. Enables optimization of decision-making at key business applications

II. ARTIFICIAL INTELLIGENCE

A. Capabilities in software to make decisions and actions without explicit instructions

B. Can leverage natural language processing

C. Enables a number of new commercial applications

III. NON-TRADITIONAL DATA

A. A number of non-traditional data sources is becoming available for financial services players

B. Enables generation of new insights across the credit value chain

IV. ADVANCED PROCESS AUTOMATION/ DIGITIZATION

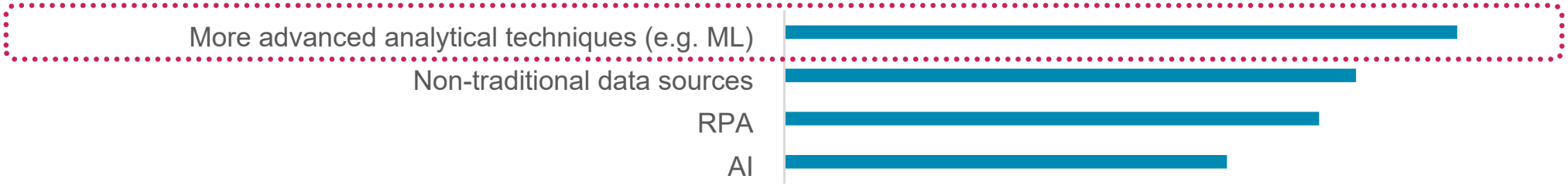
A. Automation of repeatable tasks through process digitation is ongoing

B. Combined with other enablers, enables considerable efficiency and effectiveness improvements for Risk

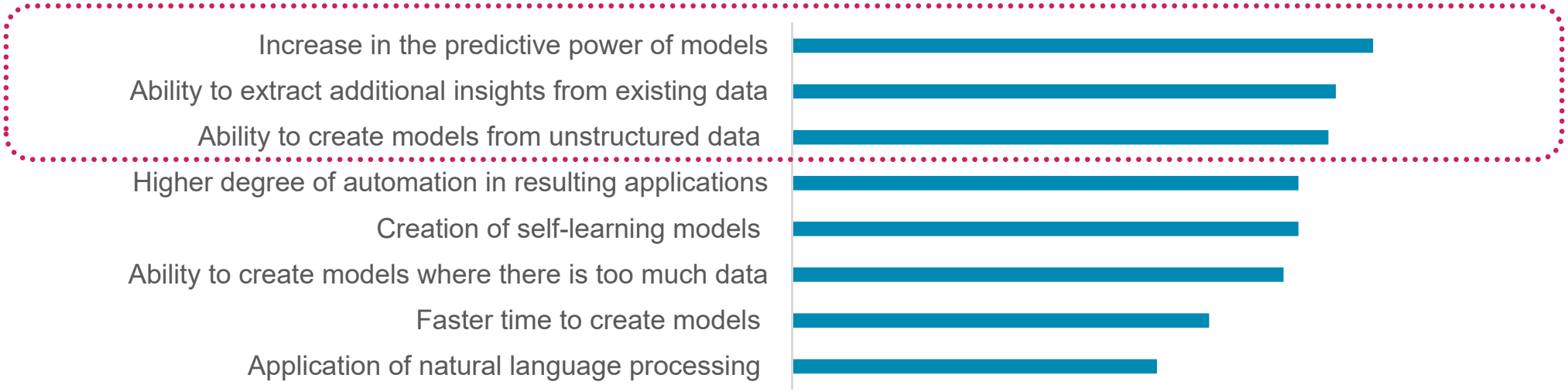
Survey results after previous Webinar (1/3)

Surveyed banks see value from more advanced analytical techniques, especially for building more predictive models

Please evaluate the mid-term (3y) business value potential of the following technology-related enables for your bank's risk department (total score given)



What would be the key value drivers that you would expect to obtain from the adoption of advanced analytics techniques? (total score given)

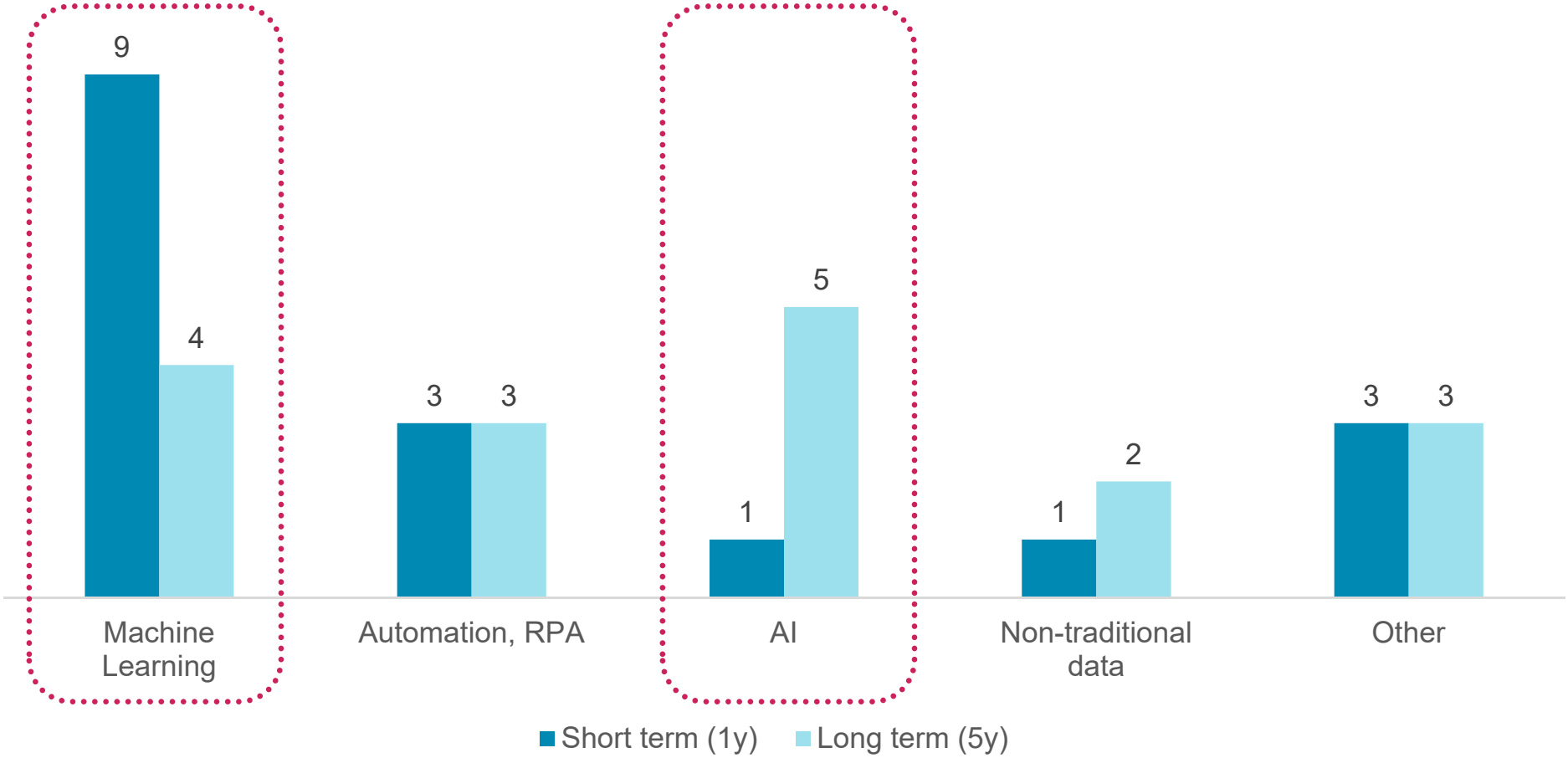


Source: Survey to participants to ABA webinar on advanced analytics on risk applications, March 2018. N = 18
Note: Total score calculated as sum of all answers where 1 = Very little or no value and 5 = Significant value)

Survey results after previous Webinar (2/3)

Surveyed banks see that Machine Learning has highest business value potential in short term, overtaken by Artificial Intelligence in long term

Please name which advanced analytics applications have the highest business value potential for your bank's risk department (number of times selected)

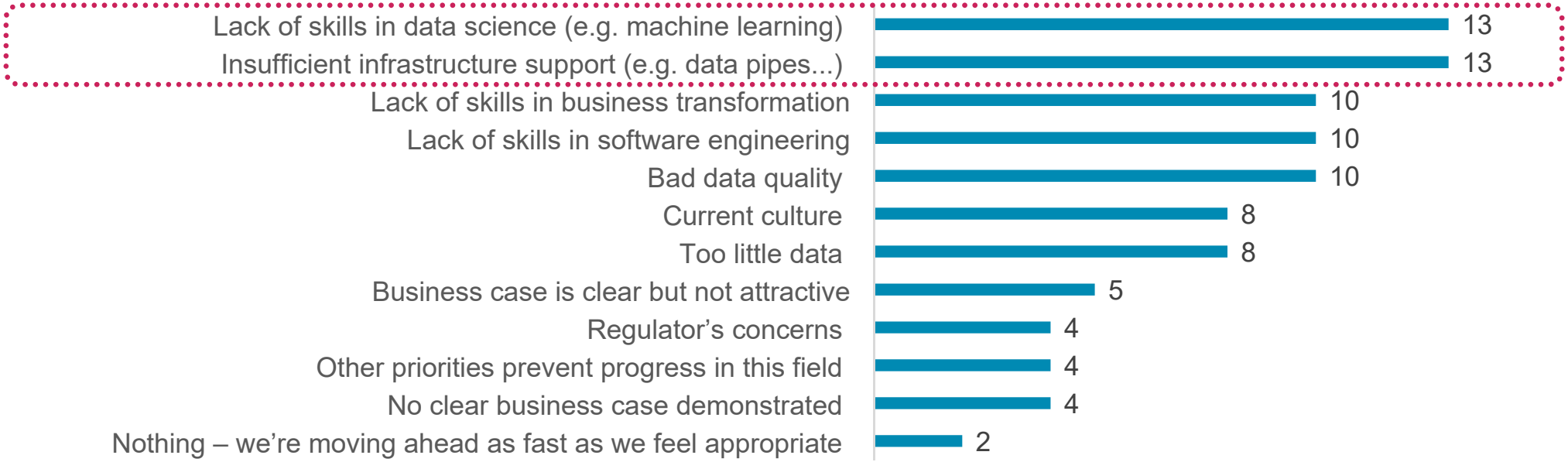


Source: Survey to participants to ABA webinar on advanced analytics on risk applications, March 2018. N = 18

Survey results after previous Webinar (3/3)

Key factors hindering progress include lack of skills in data science and insufficient infrastructure support, with also several other reasons mentioned

What's currently preventing your bank from extracting value from these technologies?
(number of times selected)



"It's issues on Change Management that we find difficult or dealing with resistance from different units to apply new technologies...."

"Lack of participation from Senior Management, Committees and the Board."

"No centralized Data Management and Data quality."

"Poor performance in model design and validation."

"Lack of expertise in model design and validation"

"Mindset"

Source: Survey to participants to ABA webinar on advanced analytics on risk applications, March 2018. N = 18

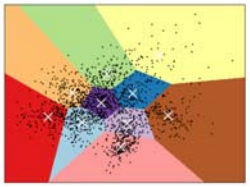
2 Credit modelling using machine learning

Credit modelling can be seen as a classification problem and thus standard supervised learning techniques are usually applicable

Machine Learning
E.g. Random Forest, Neural Network, K-means…

Unsupervised Learning Problem – "X"
Training with unlabeled data

Clustering
Algorithms group objects into clusters

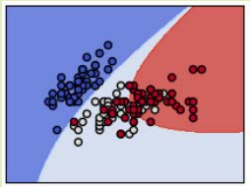


Association
People that buy X also tend to buy Y

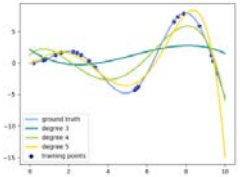
{onions, potatoes}
➔ {burger}

Supervised Learning – "Y = f(X)"
Training with known output

Classification
Output variable is the class of the object

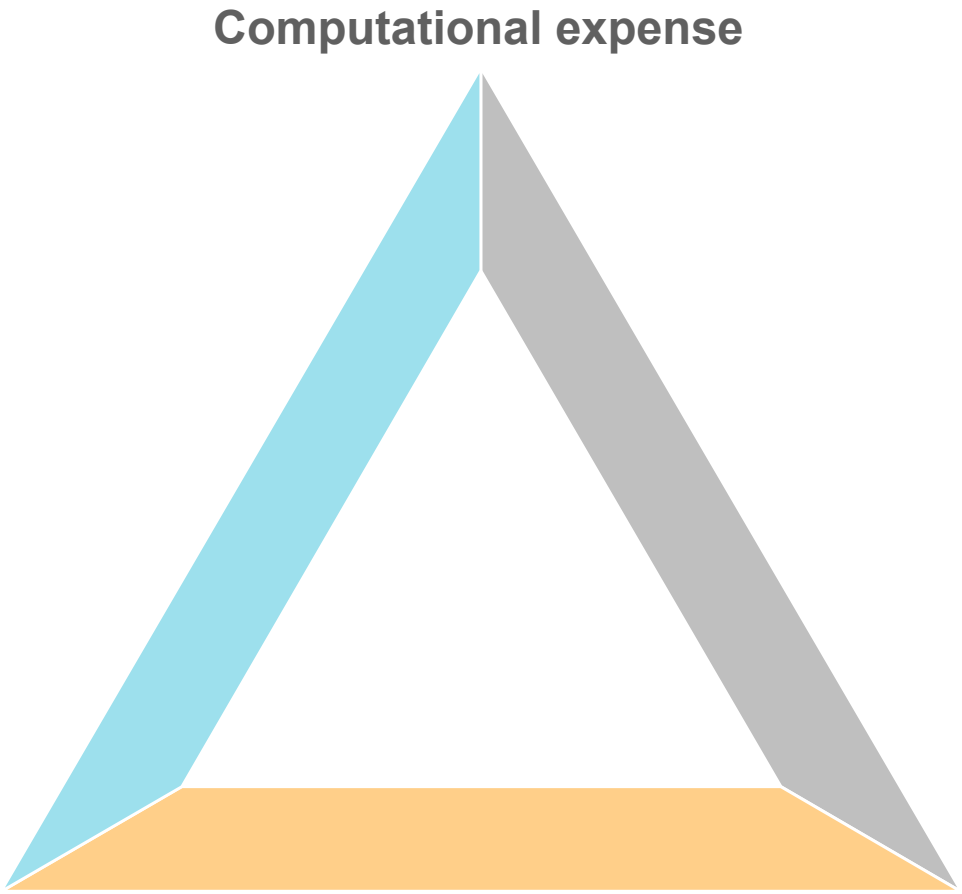


Regression
Output variable is a real value



Modelling fundamentals

Different modelling techniques strike a trade-off between predictability, interpretability, and computational expense



Predictability
Better risk differentiation;
less risk costs; more customers

Interpretability
Increased buy-in; invite challenge;
avoid overfit;

Non-exhaustive

Classification algorithms (examples)

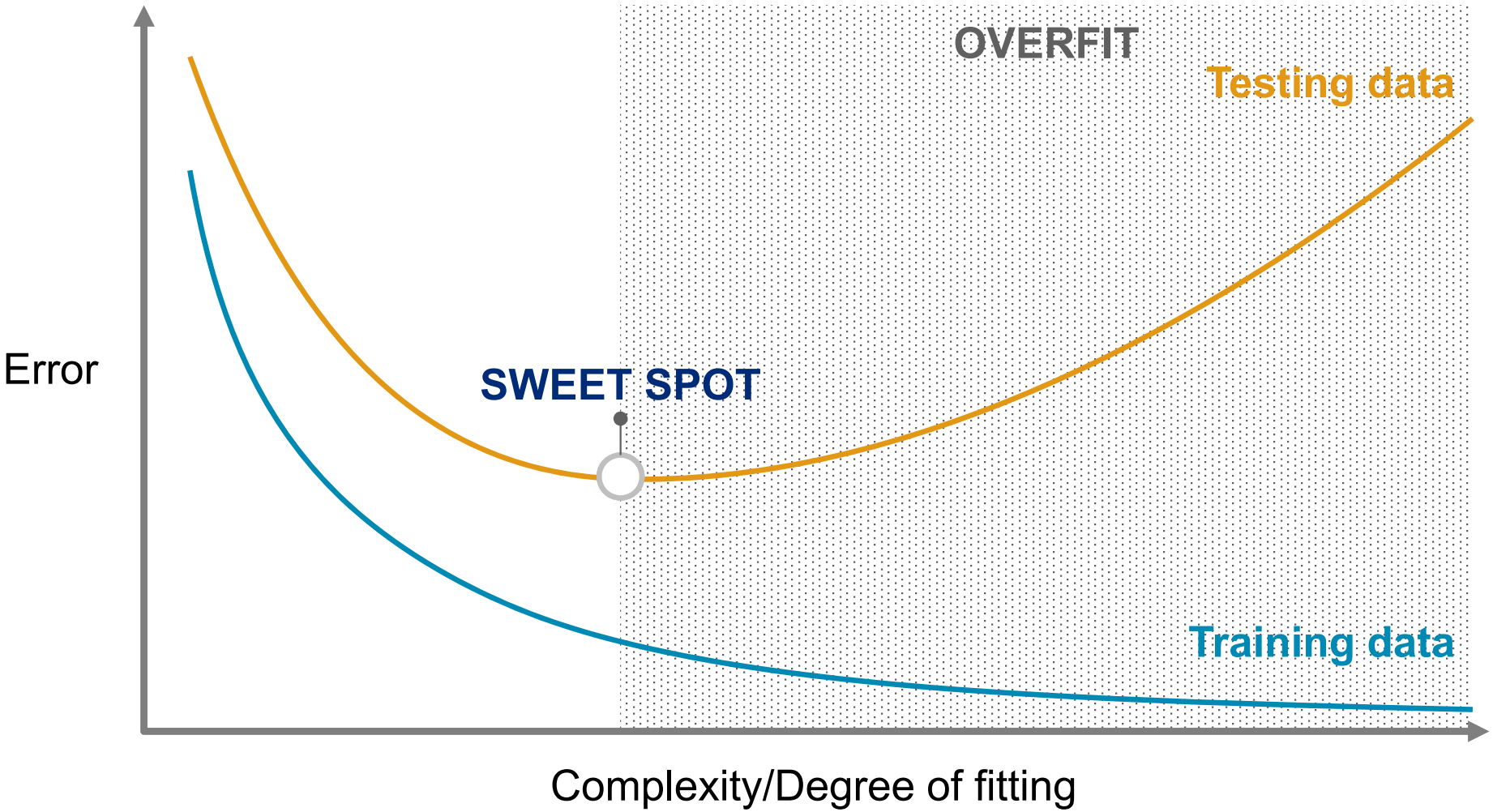
There are several Machine Learning algorithms that can be applied in classification problems with different level of interpretability

Interpretability: High → Low

Model	Computational expense
Naive Bayes	Very low
Logistic Regression	Low
Generalized Linear Models (GLM)	Low
Multivariate Adaptive Regression Splines	Low
K-Means	Low
Hierarchical clustering (Agglomerative)	Medium
Gradient-Boosted Trees	Medium
Random Forest	Medium
Support Vector Machine	High
Neural Networks	High
Multi-layer Neural Networks	Very high
Ensemble methods	High – Very high

Modelling fundamentals

Modeller needs to find a balance between minimizing prediction error and avoiding overfit to build a model that will work in application

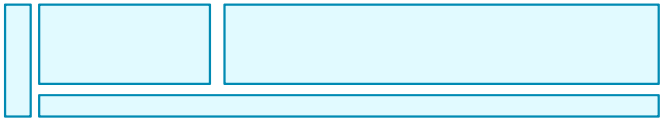


Machine Learning workflow (example)

Concentrating on building a robust machine learning workflow will help the modeller avoid overfit and find the best models

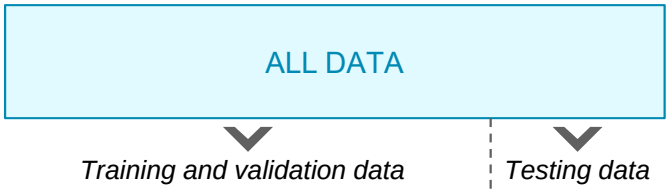
DATA ENGINEERING

- Create one data set with all data



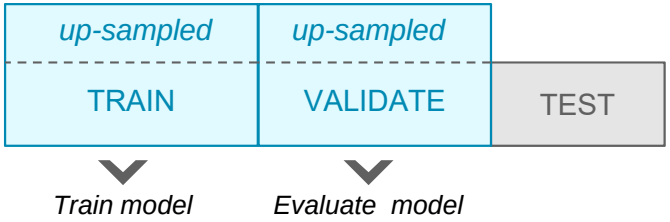
FEATURE ENGINEERING

- **Feature engineering:** Pre-processing
  - Treat missing values
  - Bucket categorical factors
  - Transform factors



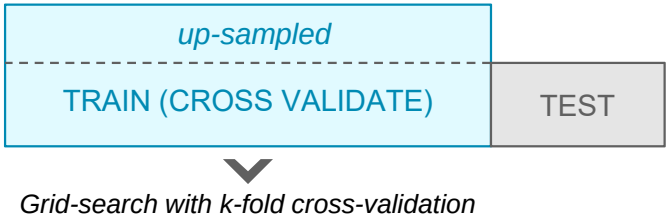
FACTOR SELECTION

- **Test classifier:** Manual factor selection based on test classifier performance (e.g. random forest), and other considerations (e.g. single-factor AUC)
- **Upsampling:** Up-sample goods to 1:1 in training data



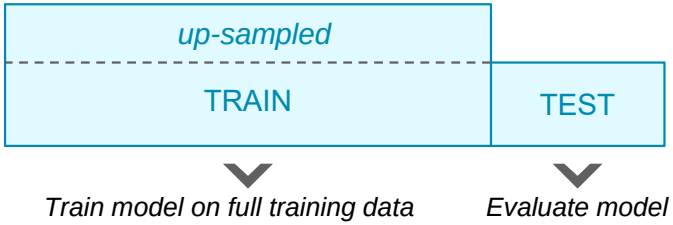
MODEL SELECTION PIPELINE

- **Grid search**: Find best model and its hyper parameters among defined set of models
- **Cross validate:** Evaluate performance by cross validating within full training data (k-fold)
- **Pipeline:** Pre-processing steps defined by model



MODEL FINALIZATION

- Train chosen model to full training data
- Evaluate final model in testing data



3 Live demo

Live modelling demo

Main techniques

python™	Programming language (free)
jupyter	Development environment (free)
scikit learn	Machine learning tools (free)

Cases

DEMO 1	Machine learning workflow incl. model selection pipeline *(public classification data set, not credit)*
DEMO 2	Object-orientated credit modelling workflow *(non-retail credit data set)*

4 Implications

Challenges for credit modelling

Credit institutions face specific challenges through the credit ecosystem when implementing machine learning algorithms

	Challenges	Example questions	Potential responses
1	**The black box problem**	• How do you trust output from an algorithm you do not understand in detail? • How to train credit officers to interpret model outputs?	• **Turn the black box into a transparent box.** Models can provide clear justification and break down the rationale for results • **In the way the model is used.** There are applications of machine learning algorithms where a black box model is acceptable
2	**The validation problem**	• How do you validate a rating system that is based on machine learning algorithms?	• **Majority of "classical" validation techniques still apply** for machine learning algorithms • The key issue is **avoiding over-fitting**; this is an area where the **validation toolkit needs to be expanded** • Modellers and validation staff **need to upgrade their skillsets**
3	**The regulatory problem**	• How do you convince a regulator that the rating system is robust and not "just works"?	• **Regulators increasingly see the benefits** of using machine learning algorithms for risk measurement • **Some regulators are more advanced** than others – in some jurisdictions machine learning scorecards are more actively discussed while in most regulators are farther behind • We expect that regulatory approvals for using machine learning algorithms for risk measurement to increase in coming years

Potential for credit modelling

Banks can reap different benefits from advanced analytics depending on the depth of adoption, starting level, and the problems being solved

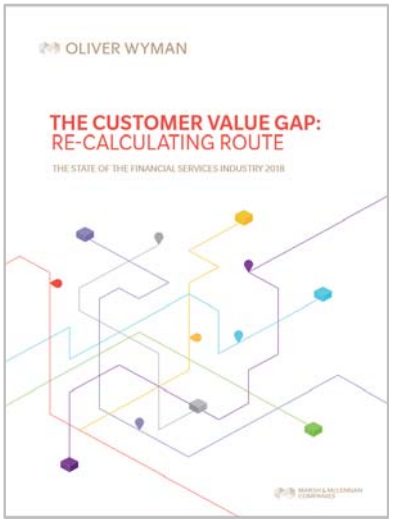
Benefits of advanced analytics toolkits in Credit Modelling

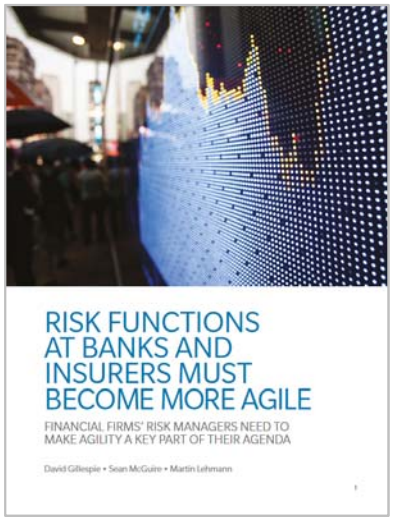
1	**Improved data processing** *"Use Python instead of old techniques"*	• More industrialized data processing, especially compared to spreadsheets – More automatic, faster, more repeatable, codified – Higher-level user involvement – Automation of documentation (e.g. Jupyter notebook) • Net impact is a shorter time to build, update, and document a model
2	**New insights** *"Use ML to gain additional insights"*	• Problems that don't require use of Machine Learning often benefit from additional insights gained from Machine Learning to data and solutions, e.g. – Calculate factor importance taking into account interactions – Discover nonlinearities – Compare performance to challenger models • Net impact is better and faster traditional models
3	**Updated models** *"Use ML as the final model"*	• Problems that have considerable nonlinearities or constitute of large data sets are often better solved by Machine Learning, e.g. – Use ANN instead of regression for retail/SME credit decisioning – Fraud detection – Big data where "old toolkit" is not powerful enough • Net impact is new models fully leveraging advanced analytics

Effort required for data preparation will continue to be considerable

Any questions?

Oliver Wyman Reports





Contact Us

For further information, follow us on Asia Pacific Risk Center (APRC) and BRINK Asia

ASIA PACIFIC RISK CENTER

www.mmc.com/asia-pacific-risk-center.html

For information about banking news, reports, conferences, Policy positions, training, meeting opportunities, and membership, visit:

ASIAN BANKERS ASSOCIATION

www.aba.org.tw

BRINK ASIA

Economy • Environment • Geopolitics • Society • Technology

"Brink Asia is a digital news platform that provides regional perspectives from leading experts on issues related to emerging risks, growth and innovation."

www.brinknews.com/asia

@BRINKAsia

contact@brinkasia.com

Follow 'BRINK Asia'

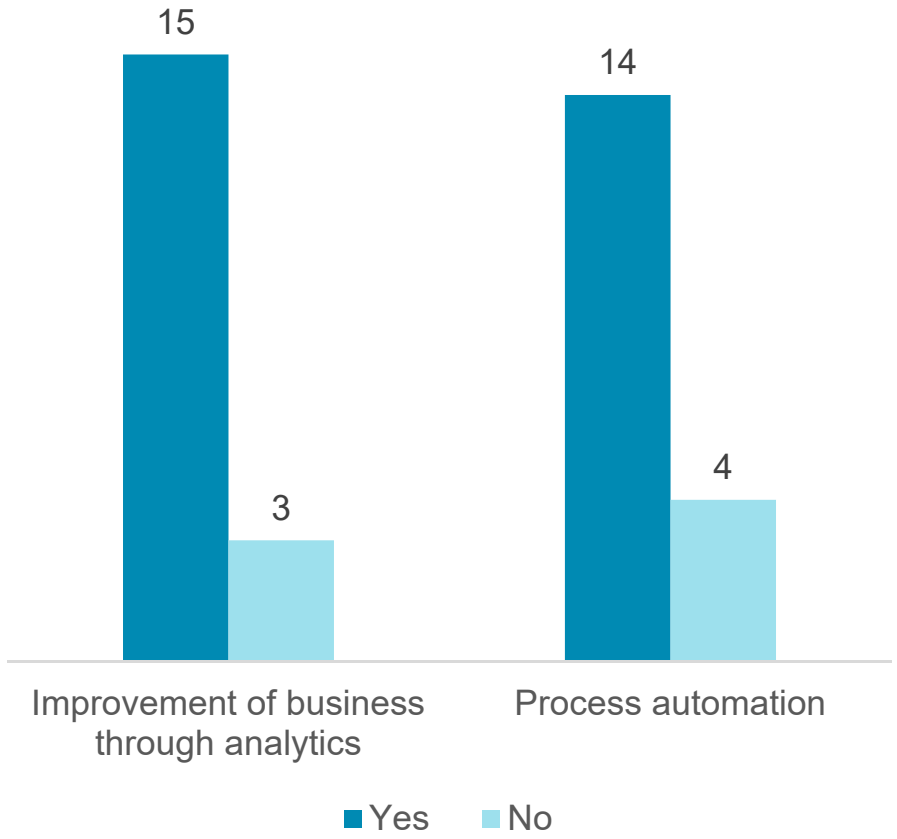
This is made possible by Marsh & McLennan Companies and managed by Atlantic Media Strategies

Appendix | Supporting pages

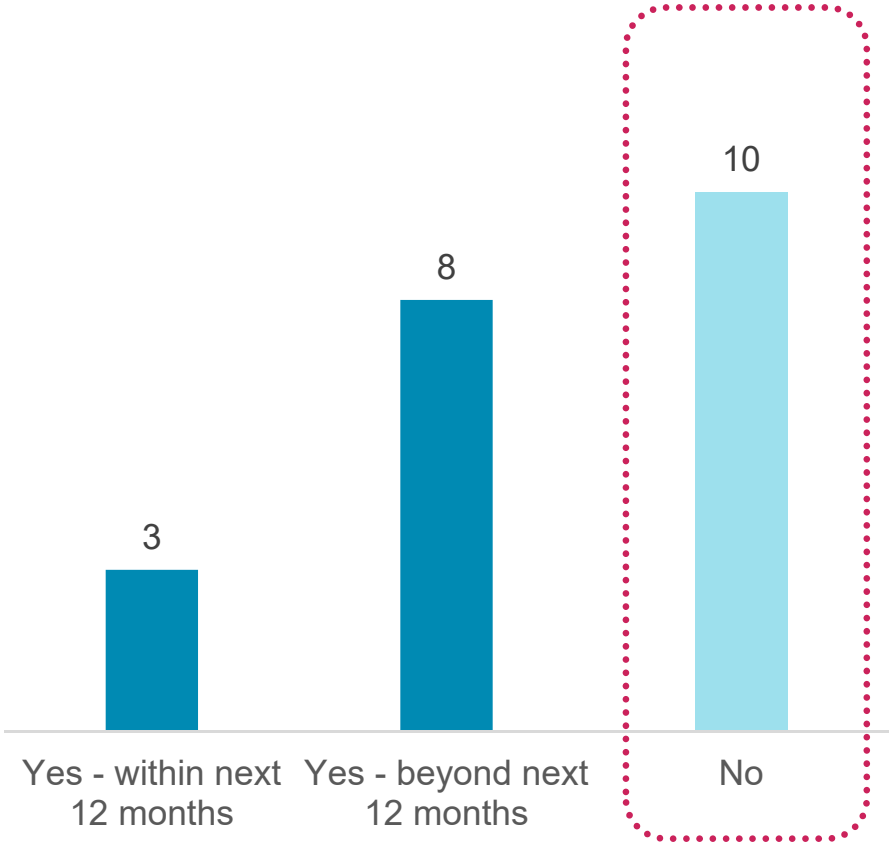
Survey results after previous Webinar

While majority of surveyed banks consider analytics and automation as strategic priorities this is not yet reflected in action plans for most of them

Are following indicated as strategic priorities for your bank (e.g. by senior management)? (number of times selected)



Does your bank have concrete action plans to build further capabilities for advanced analytics? (number of times selected)

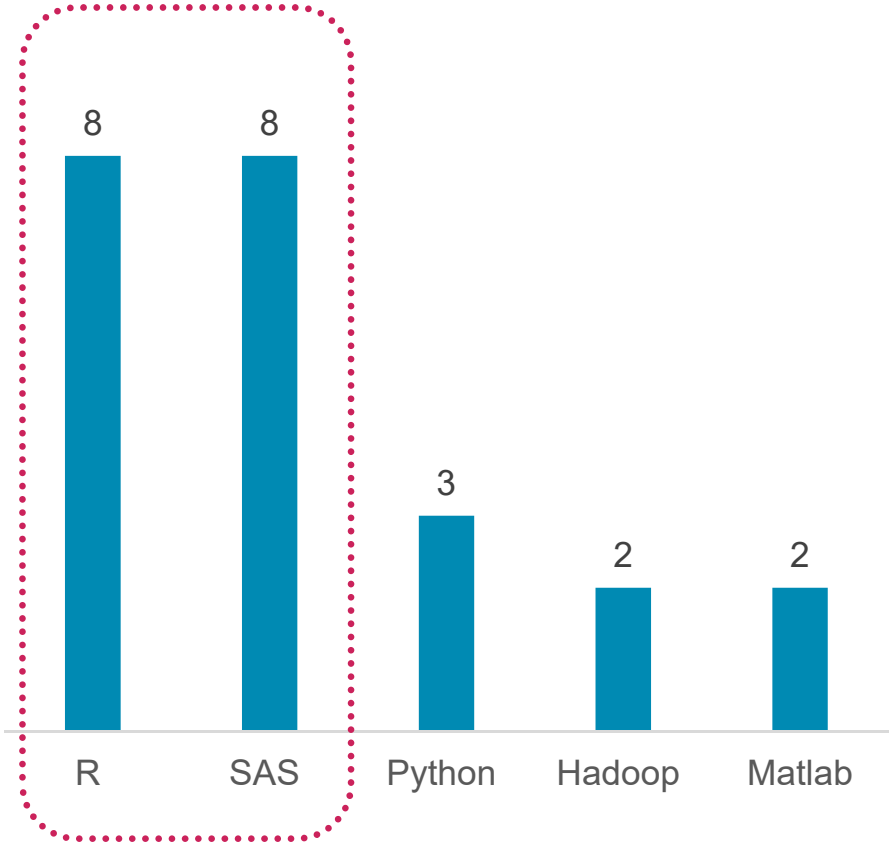


Source: Survey to participants to ABA webinar on advanced analytics on risk applications, March 2018. N = 18

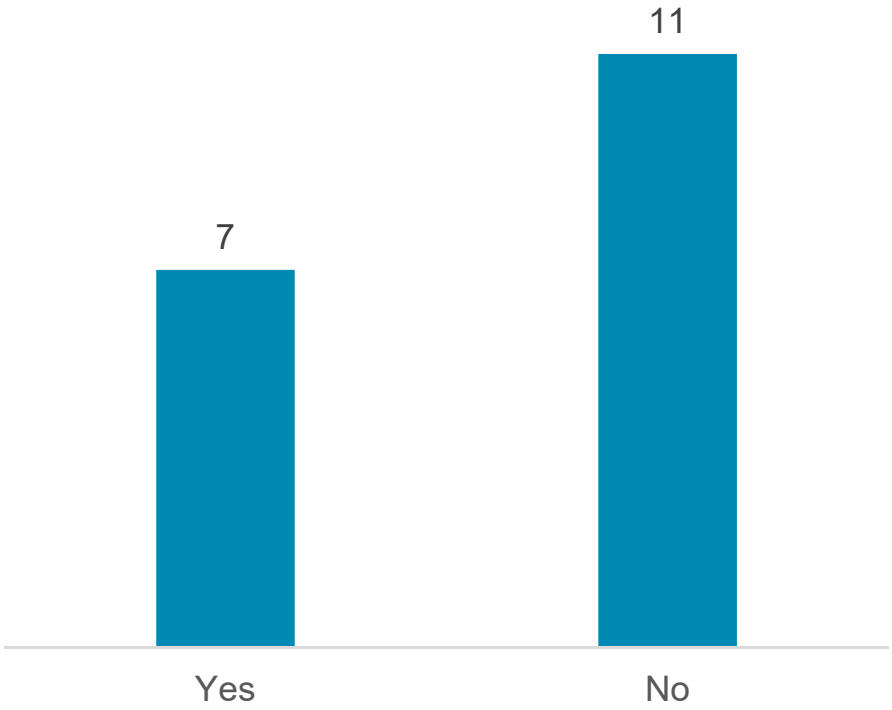
Survey results after previous Webinar

On technology side, R and SAS are identified as most common platforms planned for development and deployment of machine learning based models

Which platforms do you plan use for development and deployment of machine learning models? (number of times selected)

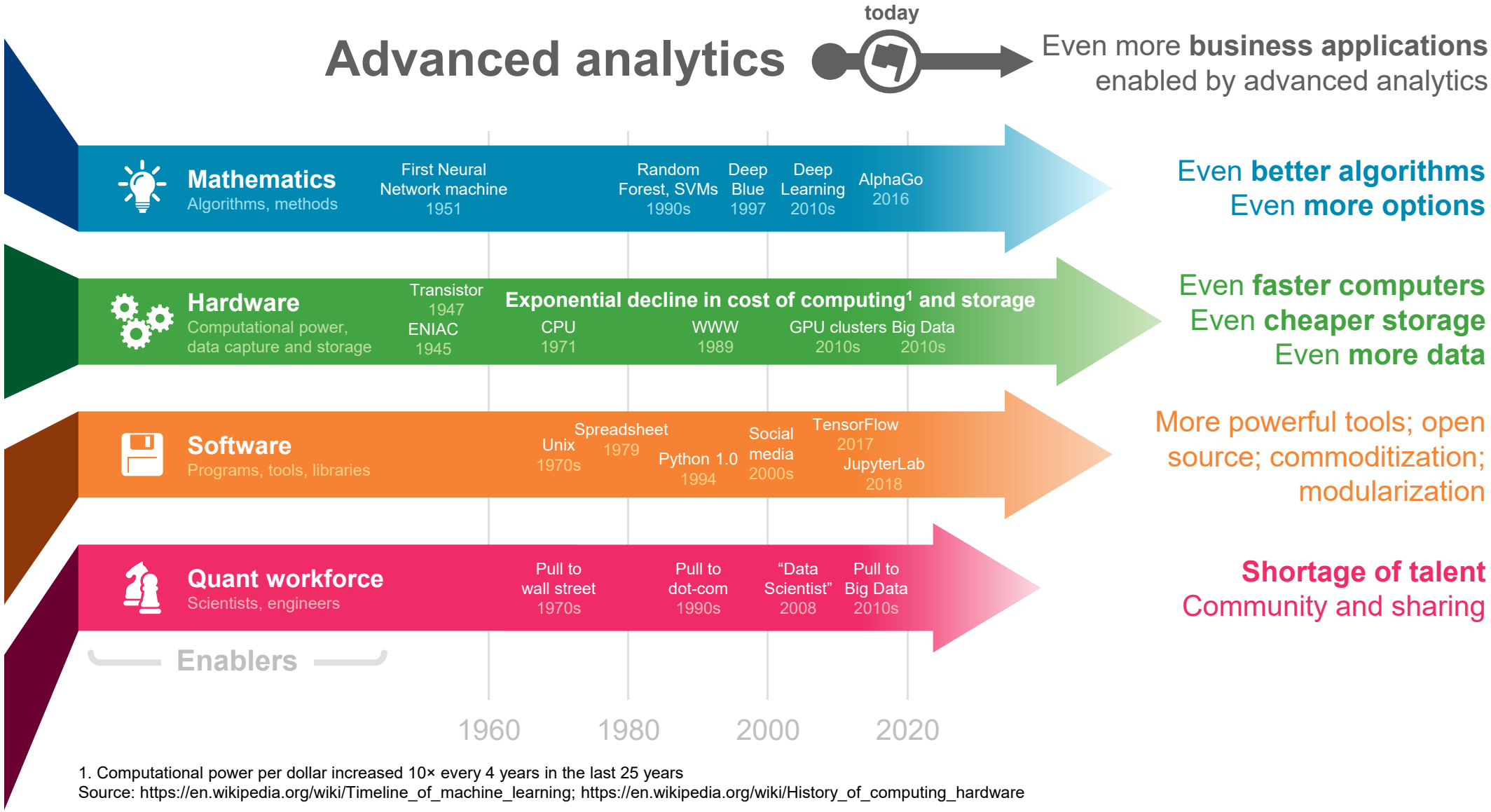


Have you developed internal capabilities to monitor and maintain machine learning based models? (number of times selected)



Source: Survey to participants to ABA webinar on advanced analytics on risk applications, March 2018. N = 18

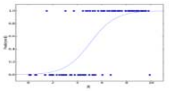
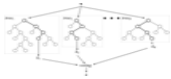
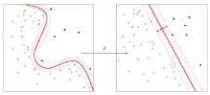
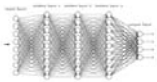
Perspective
Why Advanced Analytics now?



1. Computational power per dollar increased 10× every 4 years in the last 25 years
Source: https://en.wikipedia.org/wiki/Timeline_of_machine_learning; https://en.wikipedia.org/wiki/History_of_computing_hardware

Classification algorithms (examples)

There are several Machine Learning algorithms that can be applied in classification problems with different level of interpretability

Model	Description	When it's used	Interpretability	Computational expense
Naïve Bayes	Simple probabilistic classifier assuming feature independence	Create a baseline prediction	High	Very low
Generalized Linear Models 	Basic but powerful classifiers including logistic regression and regularized linear regression (LASSO, Ridge, Elastic Net)	Model linearly separable data (or nonlinear interactions can easily be built)	High	Low
MARS	Non-parametric regression performs variable selection and reduces overfitting	Nonlinear separation with many features	Medium	Low
Bagged Trees	Averaged decision tree classifiers trained on random subsets of data	Nonlinear separation with many features	Low	Medium
Random Forest 	Averaged decision tree classifiers trained on random subsets of data and features	Nonlinear separation with many features Unstructured data	Low	Medium
Gradient-Boosted Trees 	Simple decision tree classifiers reweighted on misclassified points	Nonlinear separation with many features	Very low	Medium
Support Vector Machines 	Separation of data in an n-dimensional space through data transformations, maximizing separation of data clusters to perform predictions	Nonlinear separation with few features and many data points	Low	High
Artificial Neural Networks 	Logistic regressions on several layers between variables and output	Few features and highly complex, nonlinear separation	Low	Very high

QUALIFICATIONS, ASSUMPTIONS AND LIMITING CONDITIONS

This report is for the exclusive use of the Oliver Wyman client named herein. This report is not intended for general circulation or publication, nor is it to be reproduced, quoted or distributed for any purpose without the prior written permission of Oliver Wyman. There are no third party beneficiaries with respect to this report, and Oliver Wyman does not accept any liability to any third party.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. Oliver Wyman accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties.