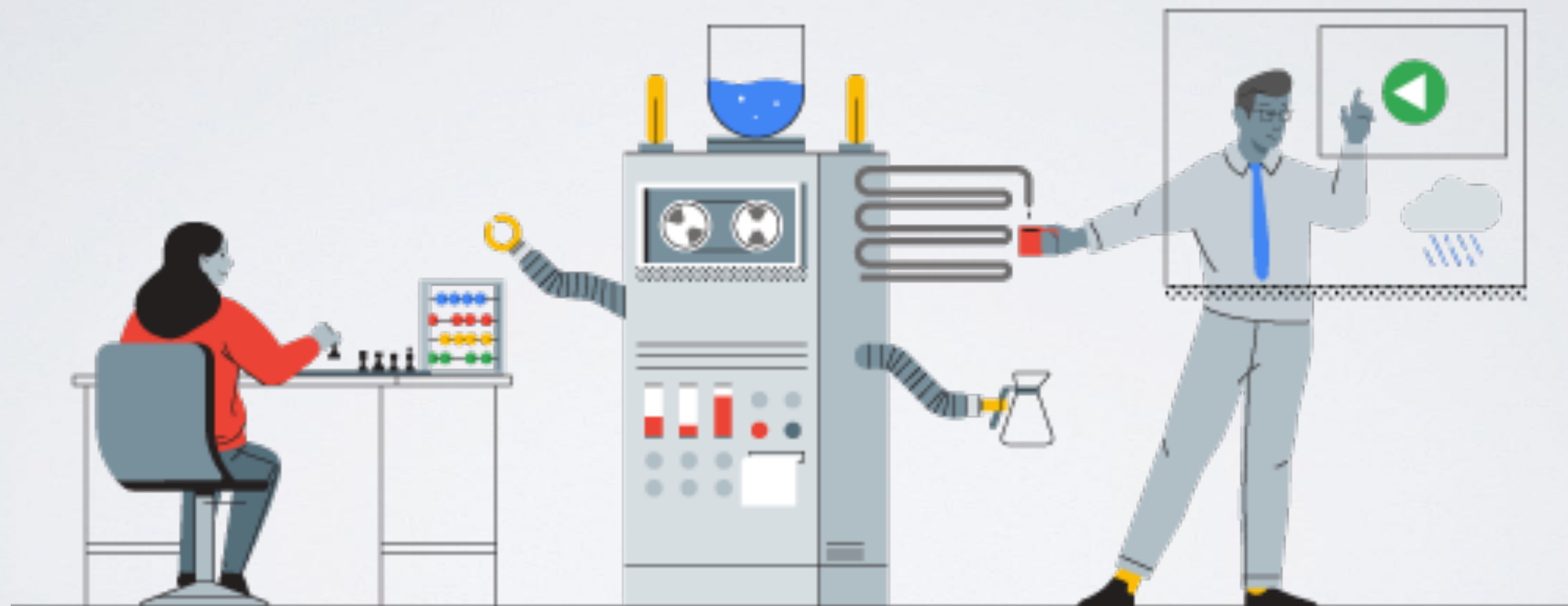
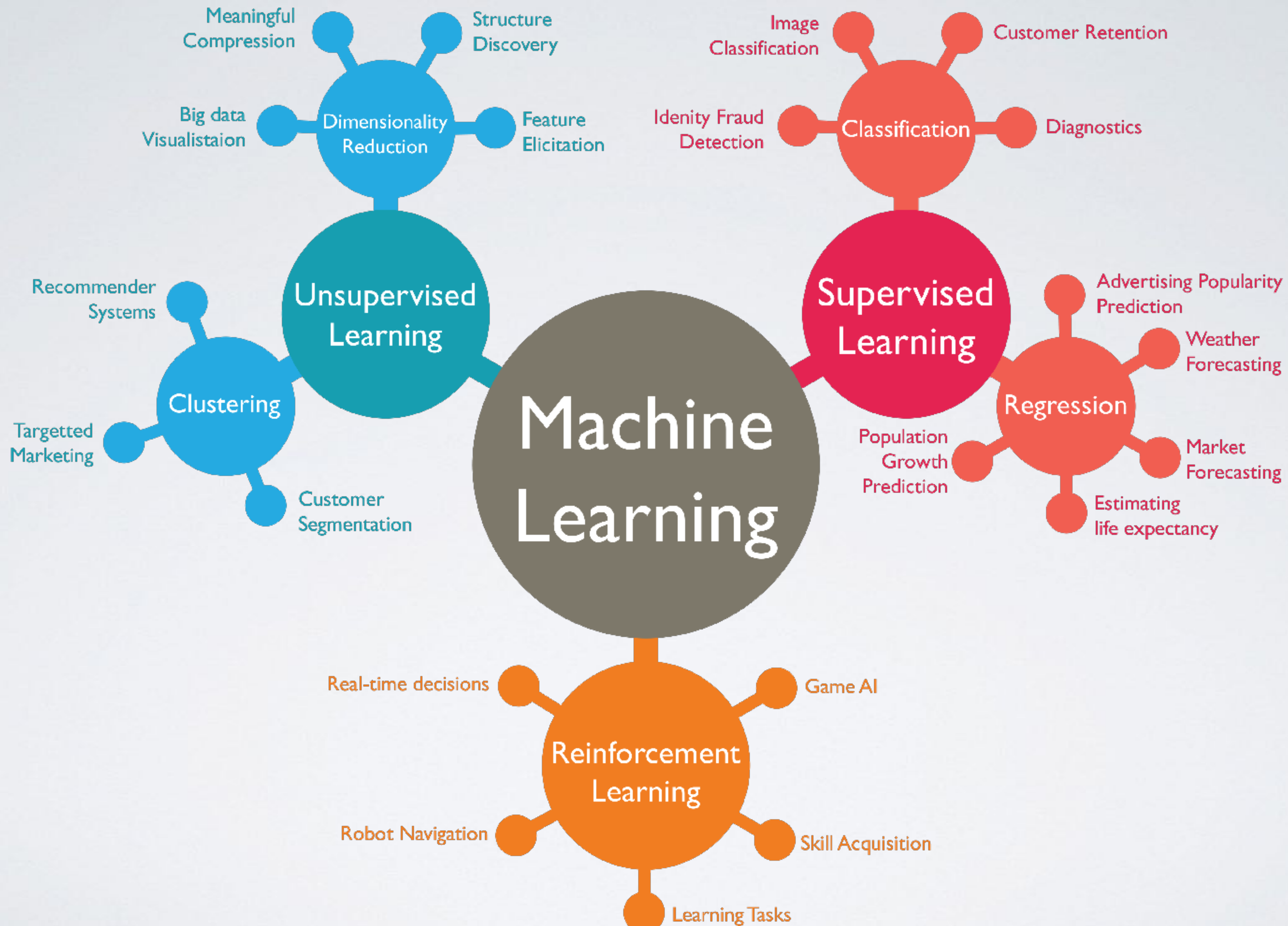


EVALUATING BUSINESS VALUE OF MACHINE LEARNING SYSTEM

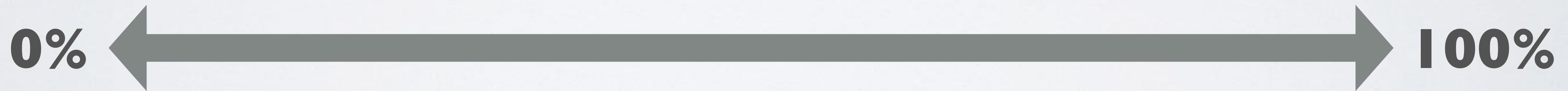
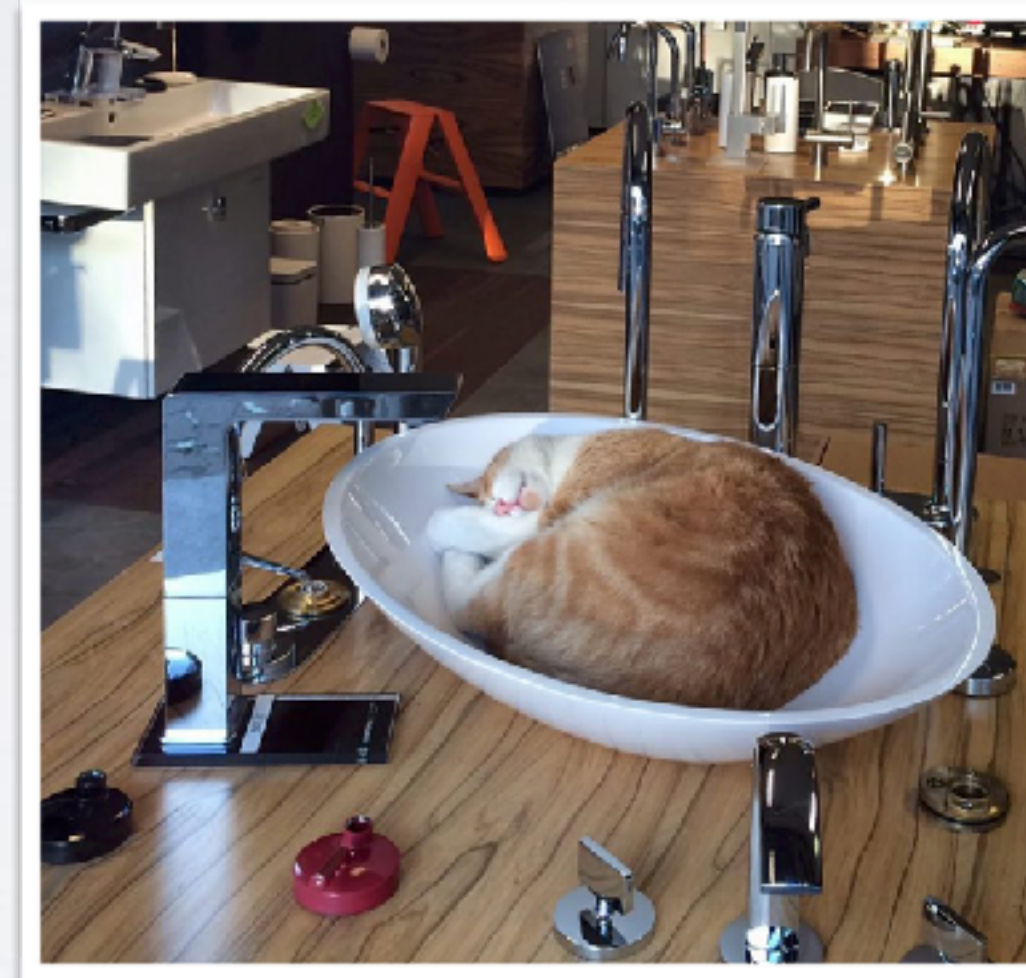
How investing into predictive
analytics can save you thousands

WHAT IS MACHINE LEARNING ?





BINARY OUTPUT VS. PROBABILITY



No, not a cat

Looks like a cat... 65%

Yes, a cat

CONFUSION MATRIX

	Is fraud	Is not fraud	
Predicted as fraud	27	4	87%
Predicted as not fraud	9	3073	99%
	75%	99%	

*Type I error
(false positive)*

*Type II error
(false negative)*

THE BOY WHO CRIED WOLF



Type I error: everyone believed there was wolf, while there was not

Type II error: everyone believed there was no wolf, while there was

SCARY SLIDE

		True condition			
Total population		Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{Condition positive}}{\Sigma \text{Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{True positive} + \Sigma \text{True negative}}{\Sigma \text{Total population}}$
Predicted condition	Predicted condition positive	True positive, Power	False positive, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{True positive}}{\Sigma \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{False positive}}{\Sigma \text{Predicted condition positive}}$
	Predicted condition negative	False negative, Type II error	True negative	False omission rate (FOR) = $\frac{\Sigma \text{False negative}}{\Sigma \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{True negative}}{\Sigma \text{Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection = $\frac{\Sigma \text{True positive}}{\Sigma \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{False positive}}{\Sigma \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{False negative}}{\Sigma \text{Condition positive}}$	True negative rate (TNR), Specificity (SPC) = $\frac{\Sigma \text{True negative}}{\Sigma \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

« In general, society places far too much value on theoretical, or academic, knowledge and not nearly enough on practical knowledge.

Knowing why planes fly does not make you a qualified pilot. »

– Nassim Nicholas Taleb
«Antifragile: Things That Gain from Disorder»

DATA TO SCORE

card_bin	issuer_bank	card_type	user_city	user_country ↑	user_ISP	sum
520373	PROMSVYAZBANK	MasterCard World Card	missing	IL	013 NetVision Ltd.	800
525477	CREDIT UNION ...	MasterCard World Card	missing	LT	TEO LT, AB	400
512762	Citibank	MasterCard World Card	Amsterdam	NL	Digital Ocean, Inc.	1500
548673	ALFABANK	MasterCard Standard ...	Amsterdam	NL	Websense Hoste...	900
546938	SBERBANK OF R...	MasterCard Standard ...	Amsterdam	NL	LeaseWeb B.V.	950
427229	VTB24	Visa Classic	Craiova	RO	Voxility S.R.L.	400
415429	ALFABANK	Visa Gold	Craiova	RO	Voxility S.R.L.	900
548673	ALFABANK	MasterCard Standard ...	Craiova	RO	Voxility S.R.L.	1400
440503	ROSBANK	Visa Classic	missing	RO	Luxoft Profession...	280
521324	TINKOFF CREDI...	MasterCard World Card	Saint Petersburg	RU	OJSC MegaFon	6185
479004	ALFABANK	Visa Platinum	Moscow	RU	OJSC MegaFon	1400
427681	SBERBANK OF R...	Electron	Irkutsk	RU	OJSC MegaFon	2900

One month, 3000 transactions, fraud rate 1%

PREDICTIONS IN NUMBERS

	Is fraud	Is not fraud	
Predicted as fraud	27	4	87%
Predicted as not fraud	9	3073	99%
	75%	99%	

PREDICTIONS IN MONEY

	Is fraud	Is not fraud	
Predicted as fraud	\$40,932	\$21,010	65%
Predicted as not fraud	\$19,425	\$1,805,446	98%
	67%	98%	

WE DID IT?

Stopped **\$61,402** by fraud detection

Earned **\$1,805,446** in good transactions

Lost **\$19,425** lost due to undetected fraud

NOT EXACTLY.

MASTER THE E-COMMERCE

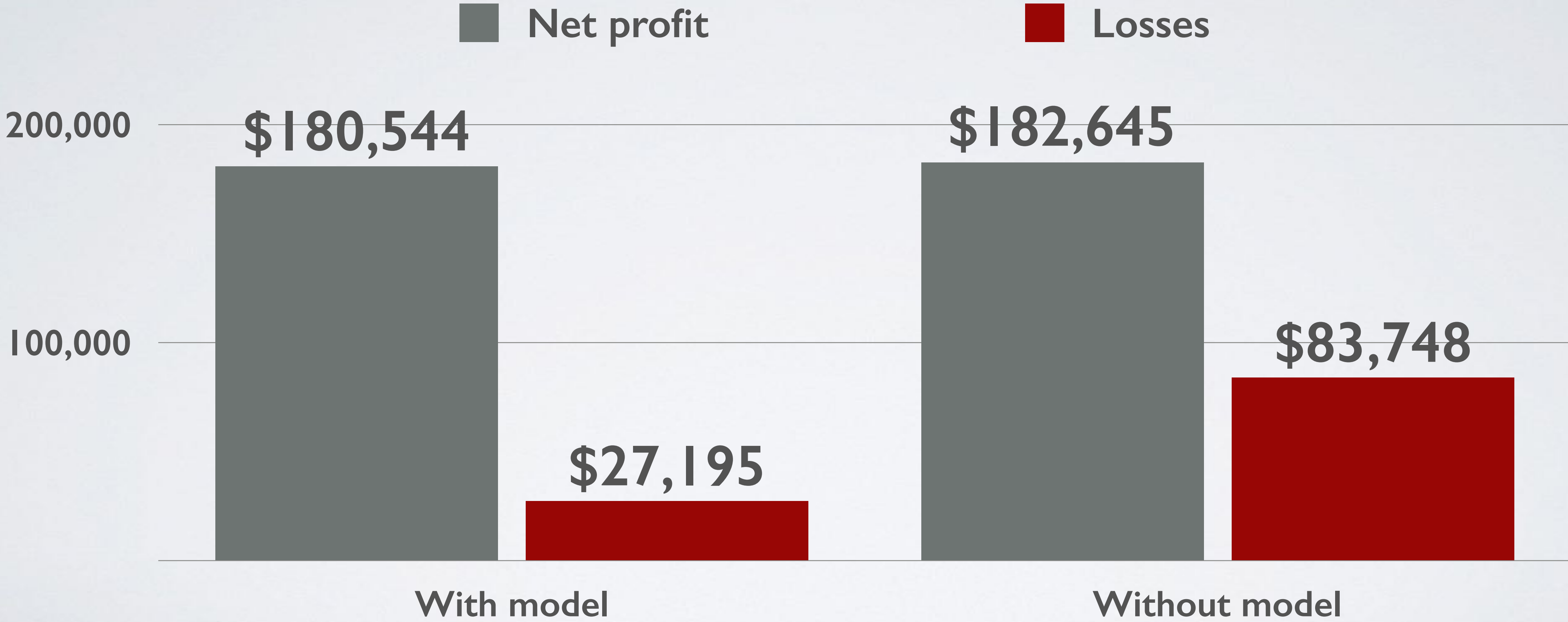
You sell for \$100, you put into your pocket only \$10!
Hence, profit margin = **10%** of revenue (turnover),
the rest is cost of running business, payroll etc

You should buy goods before you could sell them!
If fraud = \$100, you lose \$100 + average cost of
goods = \$140, or **140%** of fraudulent sum

ACTUAL MONEY FIGURES

	x1.4 ▲	x0.1 ▼
	Is fraud	Is not fraud
Predicted as fraud	\$56,548	\$2,101
Predicted as not fraud	\$27,195	\$180,544

MONEY CHART ♥



RESUME

By detecting on average 75% of fraudulent transactions, model saves us **\$56,000** per month.

Which is **\$672,000** per year.

Profit!

COST IS EVERYWHERE

E-commerce: cost of fraud, financial loss

Telecom: cost of customer churn

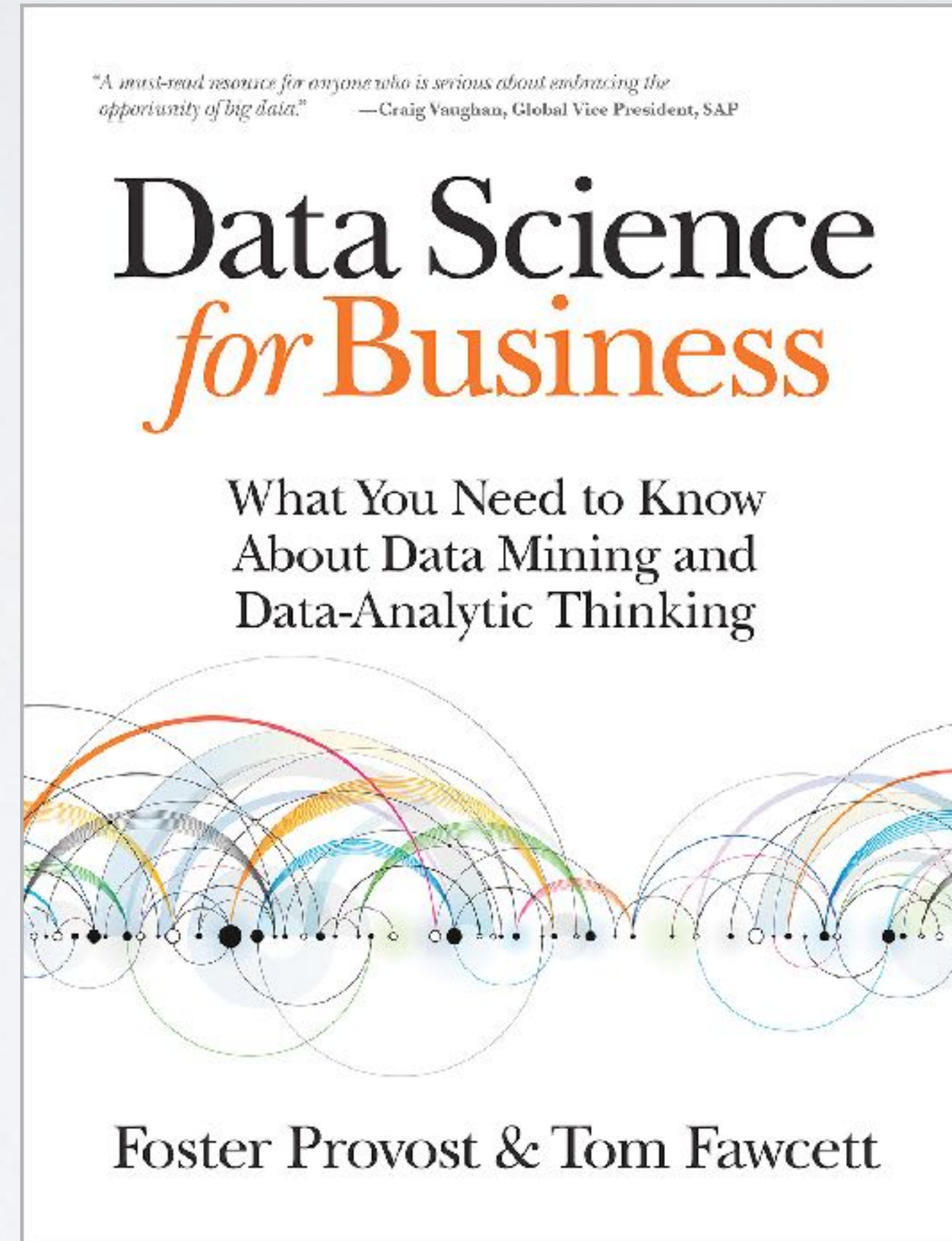
Marketing: cost of campaign

Industry: cost of component fail

Medicine: cost of wrong diagnosis

GOOD READS

the signal and the noise and the noise and the noise and the noise why so many predictions fail – but some don't and the noise and the noise and the noise silver noise noise and the noise



THANK YOU!

Vladimir Mikhnovich

WhatTheFraud.wtf

2018