



Types of Cost

Misclassification Cost

$C(i,j)$ → cost of assigning an observation of class j to class i

	Real Negative		Real Positive	
Predicted Negative	$C(0,0)$	TN	$C(0,1)$	FN
Predicted Positive	$C(1,0)$	FP	$C(1,1)$	TP

Constant Error Cost

$C(0,0)$ and $C(1,1)$ → Cost of correct classification, usually 0

$C(0,1)$ and $C(1,0)$ → Cost of FN and FP, respectively, usually 1

	Real Negative		Real Positive	
Predicted Negative	$C(0,0)$	= 0	$C(0,1)$	= 1
Predicted Positive	$C(1,0)$	= 1	$C(1,1)$	= 0



Cost ...

- How do we define cost?
- Is Cost constant?
- Is Cost the same for all the observations?
- What does Cost depend on?
- Should I care about Cost at all?



Conditional Cost

The cost of a misclassification error is conditional on the circumstances.

Cost depends on the nature of the case / observation.

- **Fraud:** the cost of missing a fraudulent application depends on the money involved in the application
- **Medical diagnosis:** the cost of misclassifying a patient depends on the patient and the disease, .e.g., more costly in elderly patients



Cost Conditional on time

The cost of a misclassification depends on timing.

- **Sensor defect detection:** different cost of missing a defect if we have 1 month till the effect occurs or if it is happening now.



Potential solutions

Expand the classification:

- Healthy / sick and young / sick and elderly
- Defect now / defect in a week / defect in a month / no defect

Cost of test or feature

- **In finance:** cost of acquiring variables from 3rd parties
- **In medicine:** cost of carrying out the tests
 - Money
 - Stress for the patient / side effects
 - Threat to patients life / wellbeing



Cost of teacher or intervention

- **In finance:** cost of fraud investigators
- **In medicine:** cost of a professional
- **Pipeline spillage:** cost of professional repair





Computational Cost

Data Storage – Model training and use

- Time
- Money
- CO2 for the environment



Data Cost

Cost associated with acquiring the data, in particular the rare class

- Cost of 3rd party labelling e.g., Amazon Turk
- Cost of our staff labelling the data
- Cost of buying data from 3rd parties, e.g., credit data from credit agencies.

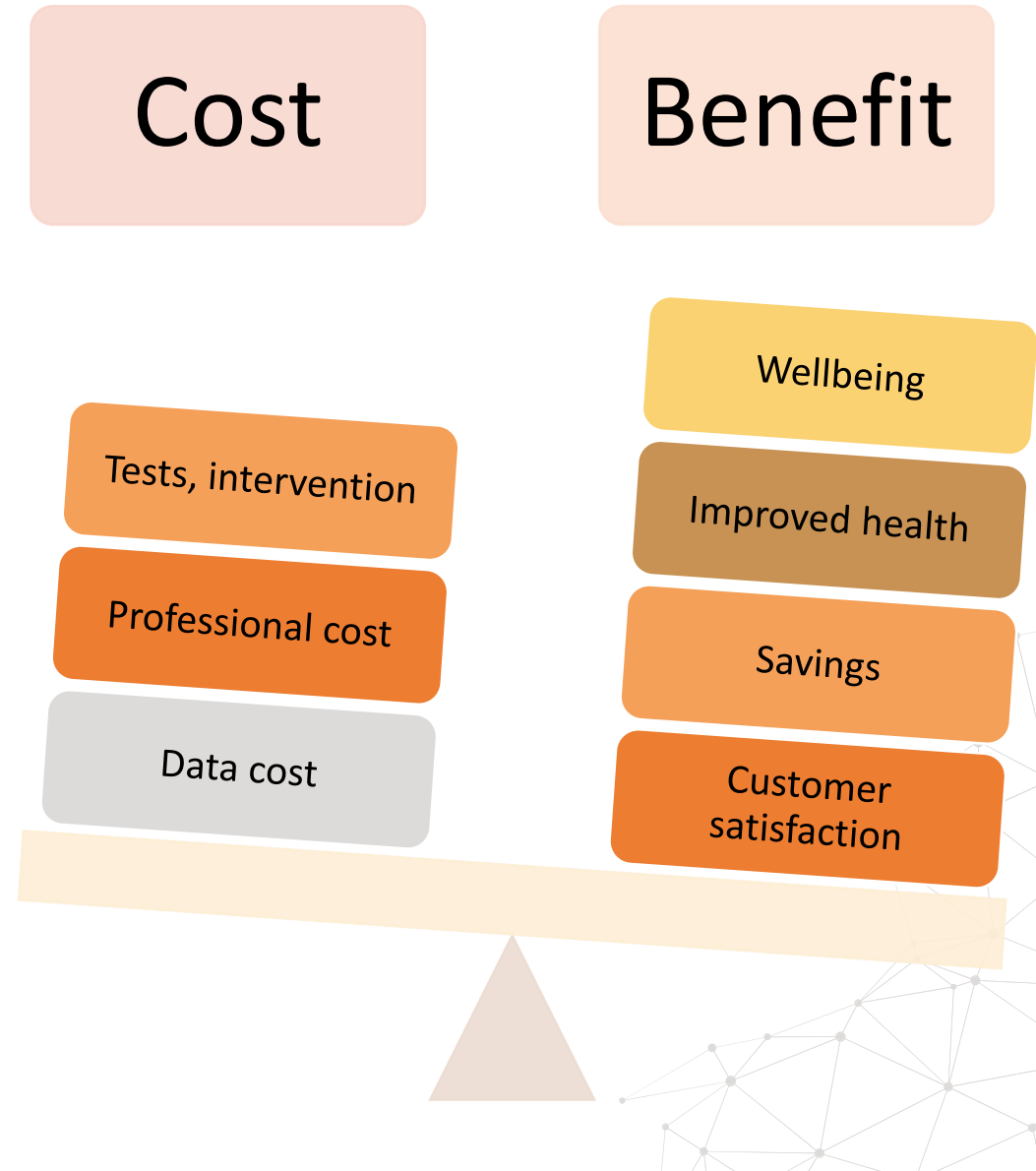
Human – Computer Cost

Cost associated with acquiring the data and building the models

- Data Analysts
- Data Engineers
- Data Scientists
- ML engineers / software developers
- Domain Experts

Cost is Complex

More than 1 player at either side of the scale



THANK YOU

www.trainindata.com