Profit Maximizing Machine Learning

Arnout Devos
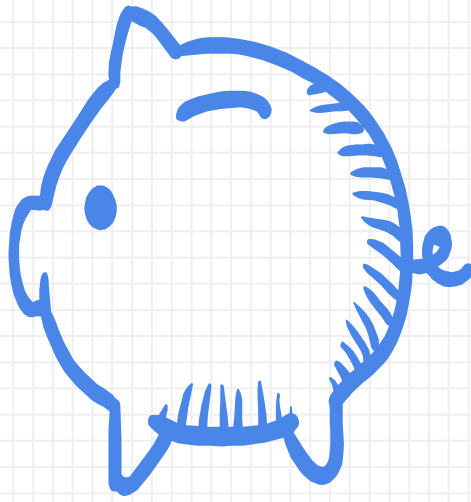
PhD student @ EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

ML is evaluated on:

Accuracy, Precision, Recall, Specificity, F1 score, Area Under the Curve, ...
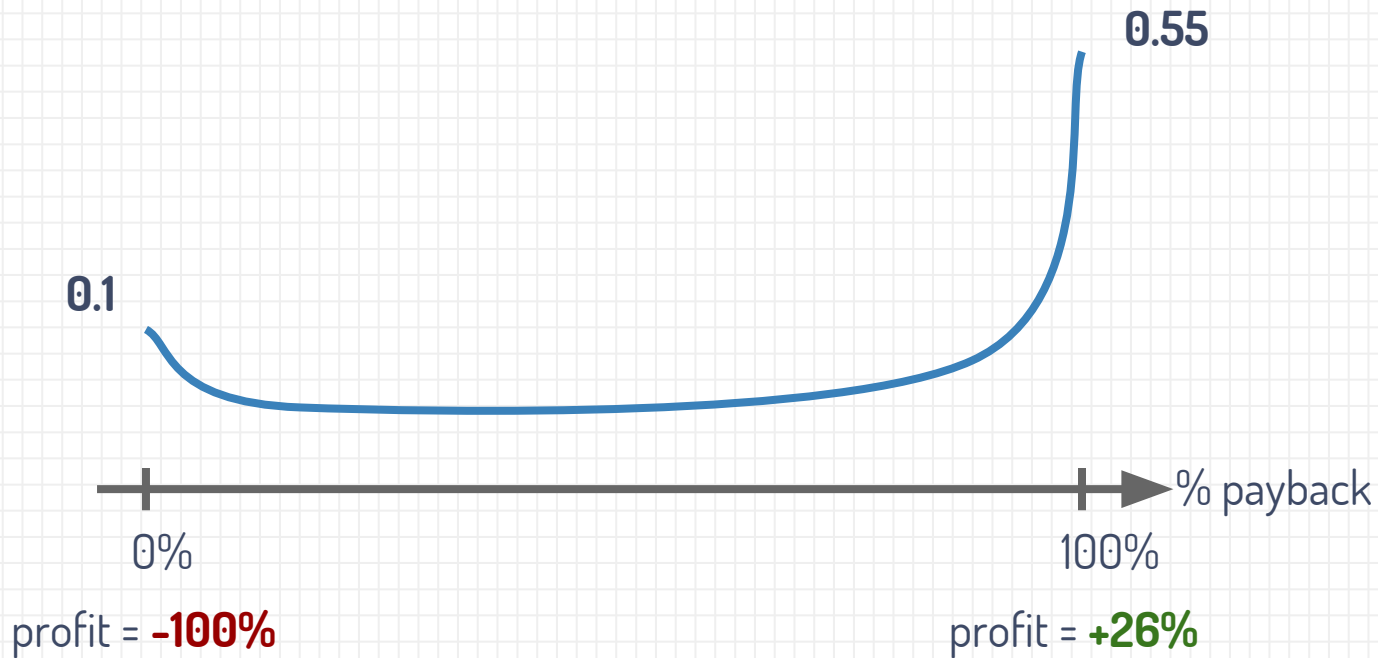
Businesses care about:

PROFIT

Maximum LIKELIHOOD credit scoring

✘ Binary goal: will applicant DEFAULT or not?
→ **Classification**

✘ Basel II/III requirements: explainable model
→ **Logistic Regression, Trees, ...**

✘ Can't use some features such as: race, gender, ...

Problem: asymmetric costs

Loan payback distribution:

0.55

0.1

% payback

0%   100%

profit = **-100%**   profit = **+26%**

Maximum PROFIT credit scoring

Combine

✗ **Logistic Regression**: explainable, easy

$$P(Y = \text{default} \mid \mathbf{x}_i) = \frac{e^{\beta_0 + \beta^T x_i}}{1 + e^{\beta_0 + \beta^T x_i}}$$

✗ **Expected Maximum Profit (EMP) measure**: profit! [Verbraken et al.], asymmetric cost problems

```
EMP = empCreditScoring(predictions,
true_classes, p0=0.55, p1=0.1, ROI=0.2644)
```

✗ **Genetic Algorithm**: (lazy) optimization

```
lr_weights = GA(lr_weights, maximize=EMP)
```

HMEQ credit dataset
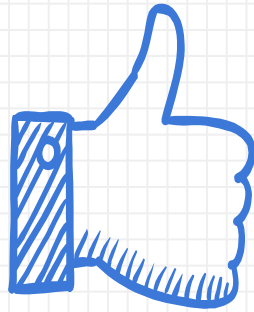
1,851,022 $
No model

2,427,683 $ +31.2 %, AUC .78
MLE Logistic Regression

2,501,595 $ +35.1 %, AUC .77
EMP Logistic Regression

THANKS!

Any questions?

You can find me at

✘   http://indy.epfl.ch

✘   @ArnoutDevos

Reference:

"Profit Maximizing Logistic Regression Modeling for Credit Scoring", Devos et al. IEEE DSW 2018