Basic churn modelling

Assumptions:

1. You have historical data about customers who churn and not churn.

2. You have a predictive model which for every customer predict whether he will churn or not churn next month.

3. You know if you give a discount $d=30\%$ to a churn customer, then with the probability $r=80\%$ he will not churn (these numbers could be assessed from historical data).

Q: I have a model of churn with accuracy of 90%. Should I use it?

Q: I have a model of churn with accuracy of 90%. Should I use it?

A: Well, not necessarily.

Q: I have a model of churn with accuracy of 90%. Should I use it?

A: Well, not necessarily.

accuracy = (TP + FP) / (TP + FP + FN + TN)

	True Churn	True not Churn
Predicted Churn	True Positive	False Positive
Predicted not Churn	False Negative	True Negative

Q: I have a model of churn with accuracy of 90%. Should I use it?

A: Well, not necessarily.

accuracy = (TP + FP) / (TP + FP + FN + TN)

	True Churn	True not Churn
Predicted Churn	True Positive	False Positive
Predicted not Churn	False Negative	True Negative

Additional assumptions:

- the fact of giving a discount will not affect False Positive customers decision;
- all customers give us the same profit **p**

Q: I have a model of churn with accuracy of 90%. Should I use it?

A: Well, not necessarily.

accuracy = (TP + FP) / (TP + FP + FN + TN)

Additional assumptions:

- the fact of giving a discount will not affect False Positive customers decision;
- all customers give us the same profit **p**

	True Churn	True not Churn
Predicted Churn	True Positive	False Positive
Predicted not Churn	False Negative	True Negative

Let's compare Total profit of doing nothing, with Total profit in case of discount strategy.

No discount (do nothing) strategy: $[FP + TN] \cdot p$

Discount strategy: $[TN + FP \cdot (1\text{-}d) + TP \cdot r \cdot (1\text{-}d)] \cdot p$

Where r is acceptance rate and d is discount

	True Churn	True not Churn
Predicted Churn	True Positive	False Positive
Predicted not Churn	False Negative	True Negative

No discount (do nothing) strategy < Discount strategy

$$[FP + TN] \cdot p < [TN + FP \cdot (1\text{-}d) \ + TP \cdot r \cdot (1\text{-}d)] \cdot p$$

$$FP + TN < TN + FP \cdot (1\text{-}d) \ + TP \cdot r \cdot (1\text{-}d)$$

$$FP < FP \cdot (1\text{-}d) \ + TP \cdot r \cdot (1\text{-}d)$$

$$FP \cdot d < TP \cdot r \cdot (1\text{-}d)$$

No discount (do nothing) strategy < Discount strategy

$$FP \cdot d < TP \cdot r \cdot (1\text{-}d)$$

$$FP \cdot 0.3 < TP \cdot 0.8 \cdot 0.7$$

$$FP \cdot 0.3 < TP \cdot 0.48$$

$$10 \cdot 0.3 \;\mathbf{?}\; 5 \cdot 0.48$$

$$3 \mathbf{>} 2.4$$

	True Churn	True not Churn
Predicted Churn	TP, 5	FP, 10
Predicted not Churn	FN, 0	TN, 85

accuracy = (TP + FP) / (TP + FP + FN + TN) = 85 + 5 / 100 = 0.9 = 90%

Q: I have a model of churn with accuracy of 90%. Should I use it?

A: Well, not necessarily.

1. Before you implement an accurate ML algorithm always compute business metrics
2. Accuracy is not appropriate for the case of imbalance classes, use ROC AUC or PR AUC instead.

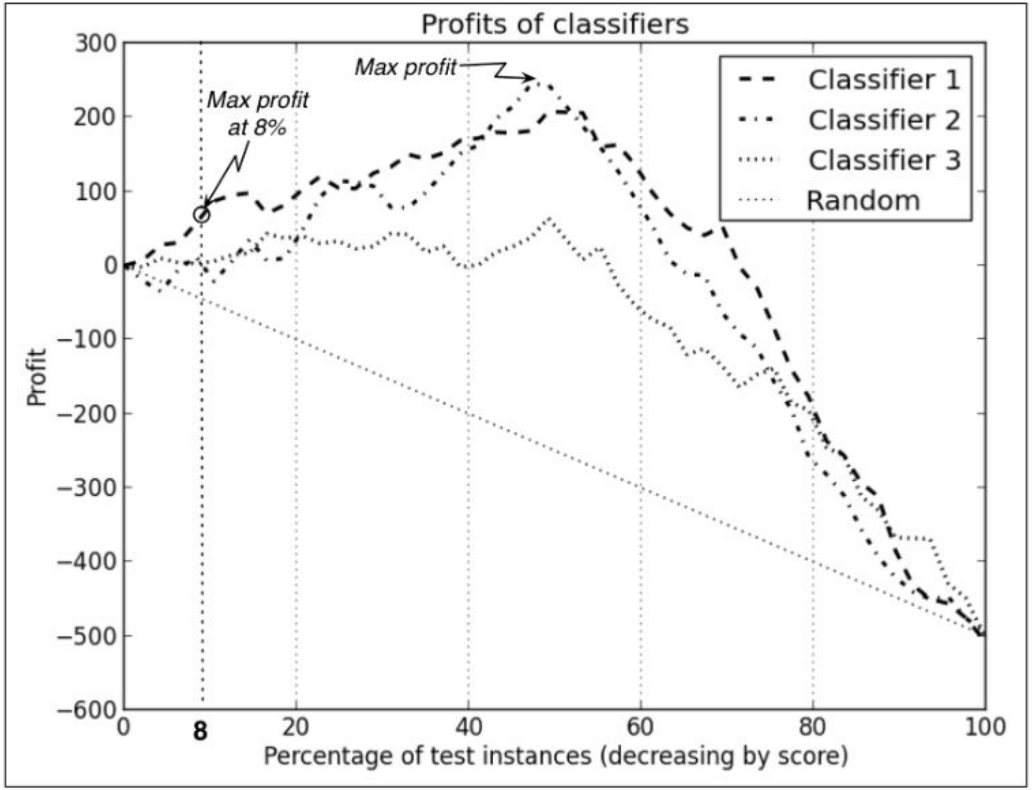

Figure 8-2. Profit curves of three classifiers. Each curve shows the expected cumulative profit for that classifier as progressively larger proportions of the consumer base are targeted.

F. Provost, T. Fawcett, Data Science for Business

Churn modelling cons

1. The fact of giving a discount will not affect False Positive customers decision.
2. We predict Churn instead of directly predicting the effect of discount.

Churn modelling cons

1. The fact of giving a discount will not affect False Positive customers decision.
2. We predict Churn instead of directly predicting the effect of discount.

Solution: Uplift

Uplift modelling

W = 1 = Treatment (e.g. discount)
W = 0 = No treatment

Train

$$fit\left(\begin{matrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{matrix} \quad \begin{matrix} w_1 \\ \dots \\ w_n \end{matrix}, \quad \begin{matrix} y_1 \\ \dots \\ y_n \end{matrix}\right)$$

$$X_{train} \qquad \boxed{W_{train}} \qquad Y_{train}$$

Prediction

$$\begin{matrix} predict \\ proba \end{matrix}\begin{pmatrix} x_{11} & \cdots & x_{1k} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mk} & 1 \end{pmatrix} - \begin{matrix} predict \\ proba \end{matrix}\begin{pmatrix} x_{11} & \cdots & x_{1k} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mk} & 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}$$

$$X_{test} \qquad W_1 \qquad\qquad X_{test} \qquad W_0 \qquad uplift$$