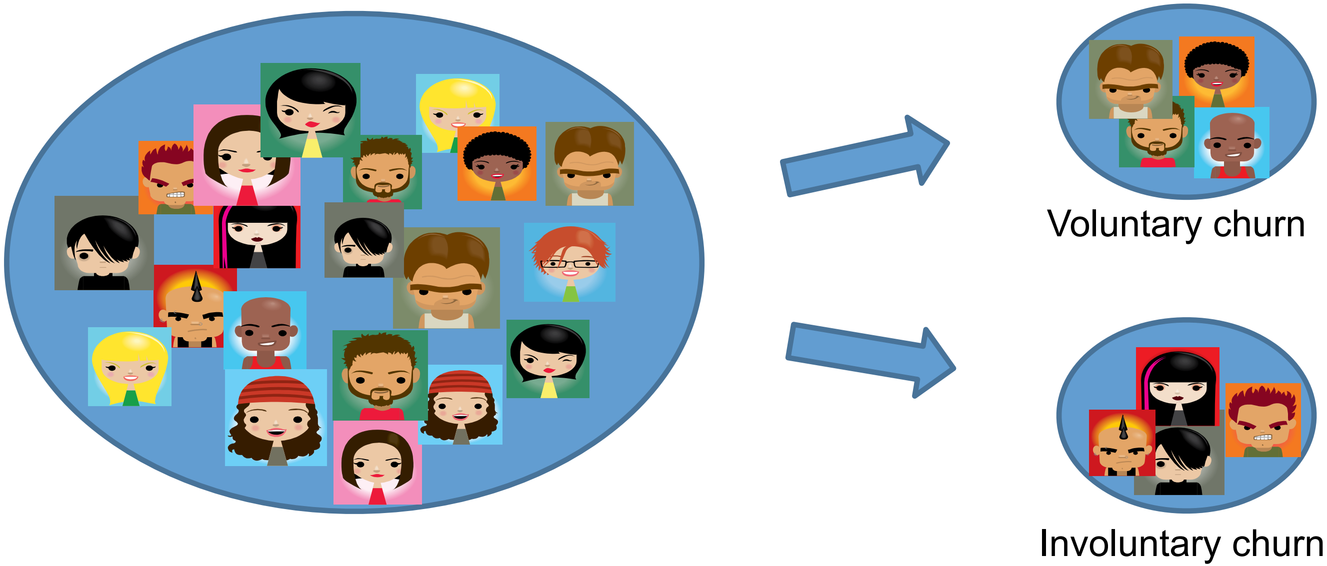
Maximizing a churn campaign's profitability with cost-sensitive Machine Learning

Agenda

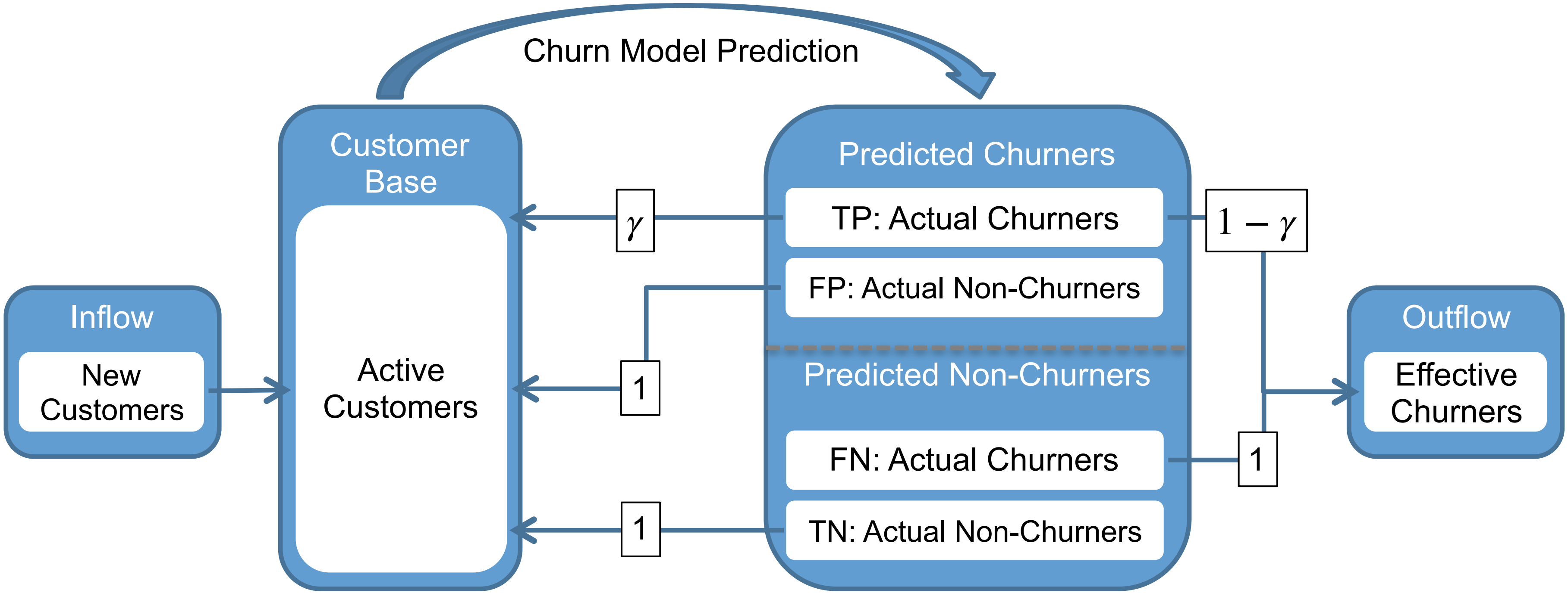
- Churn modeling
- Evaluation Measures
- Offers
- Predictive modeling
- Cost-Sensitive Predictive Modeling
    - Cost Proportionate Sampling
    - Bayes Minimum Risk
    - CS – Decision Trees
- Conclusions

Churn modeling

- Detect which customers are likely to abandon



Customer Churn Campaign Modeling



*Verbraken et. al (2013). A novel profit maximizing metric for measuring classification performance of customer churn prediction models.

Evaluation of a Campaign

- Confusion Matrix

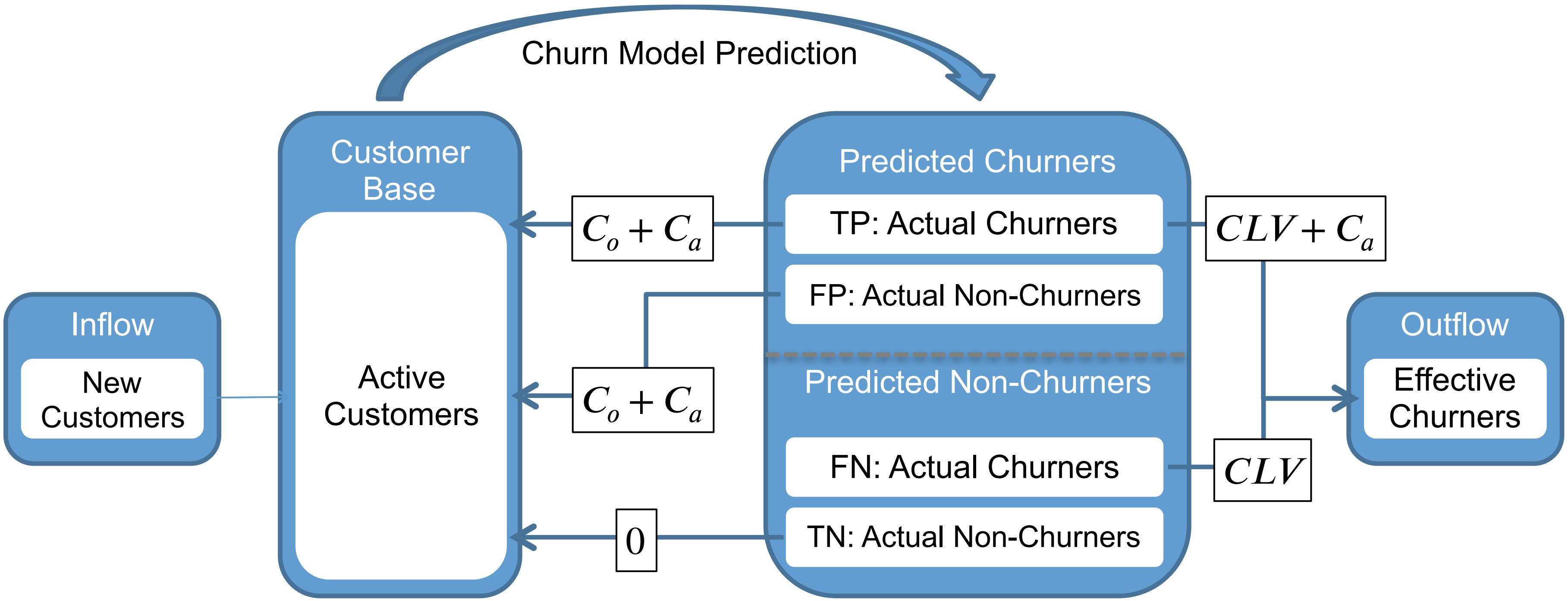
		True Class ()	
		Churner (=1)	Non-Churner(=0)
Predicted class ()	Churner (=1)	TP	FP
	Non-Churner (=0)	FN	TN

- Accuracy $= \dfrac{TP + TN}{TP + TN + FP + FN}$
- Recall $= \dfrac{TP}{TP + FN}$
- Precision $= \dfrac{TP}{TP + FP}$
- F1-Score $= 2\dfrac{Precision \ * \ Recall}{Precision + Recall}$

Evaluation of a Campaign

- However these measures assign the same weight to different errors
- Not the case in a Churn model since
  - Failing to predict a churner carries a different cost than wrongly predicting a non-churner
  - Churners have different financial impact

Financial Evaluation of a Campaign



Financial Evaluation of a Campaign

- Cost Matrix

		True Class ()	
		Churner (=1)	Non-Churner(=0)
Predicted class ()	Churner (=1)	$C_{TP_i} = \gamma_i C_{o_i} + (1-\gamma_i)CLV_i + C_a$	$C_{FP_i} = C_{o_i} + C_a$
	Non-Churner (=0)	$C_{FN_i} = CLV_i$	$C_{TN_i} = 0$

where:

Administrative cost

Cost of the offer made to customer

Client Lifetime Value of customer

Probability that customer accepts the offer

Financial Evaluation of a Campaign

- Using the cost matrix the total cost is calculated as:

$$C = \sum \left\{ y_i \left(c_i \bullet C_{TP_i} + (1 - c_i) C_{FN_i} \right) + (1 - y_i) \left(c_i \bullet C_{FP_i} + (1 - c_i) C_{TN_i} \right) \right\}$$

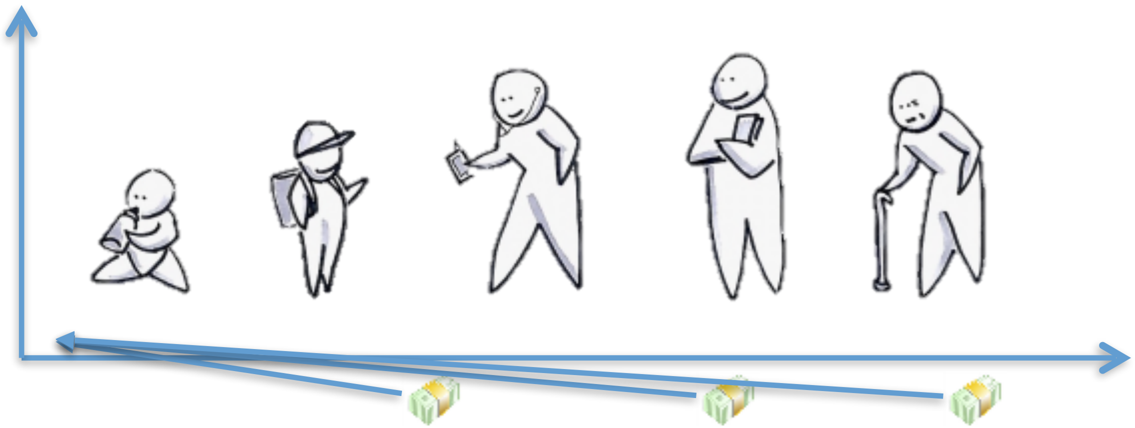
- Additionally the savings are defined as:

$$C_s = \frac{C_0 - C}{C_0}$$

where C_0 is the cost when all the customers are predicted as non-churners

Financial Evaluation of a Campaign

- Customer Lifetime Value



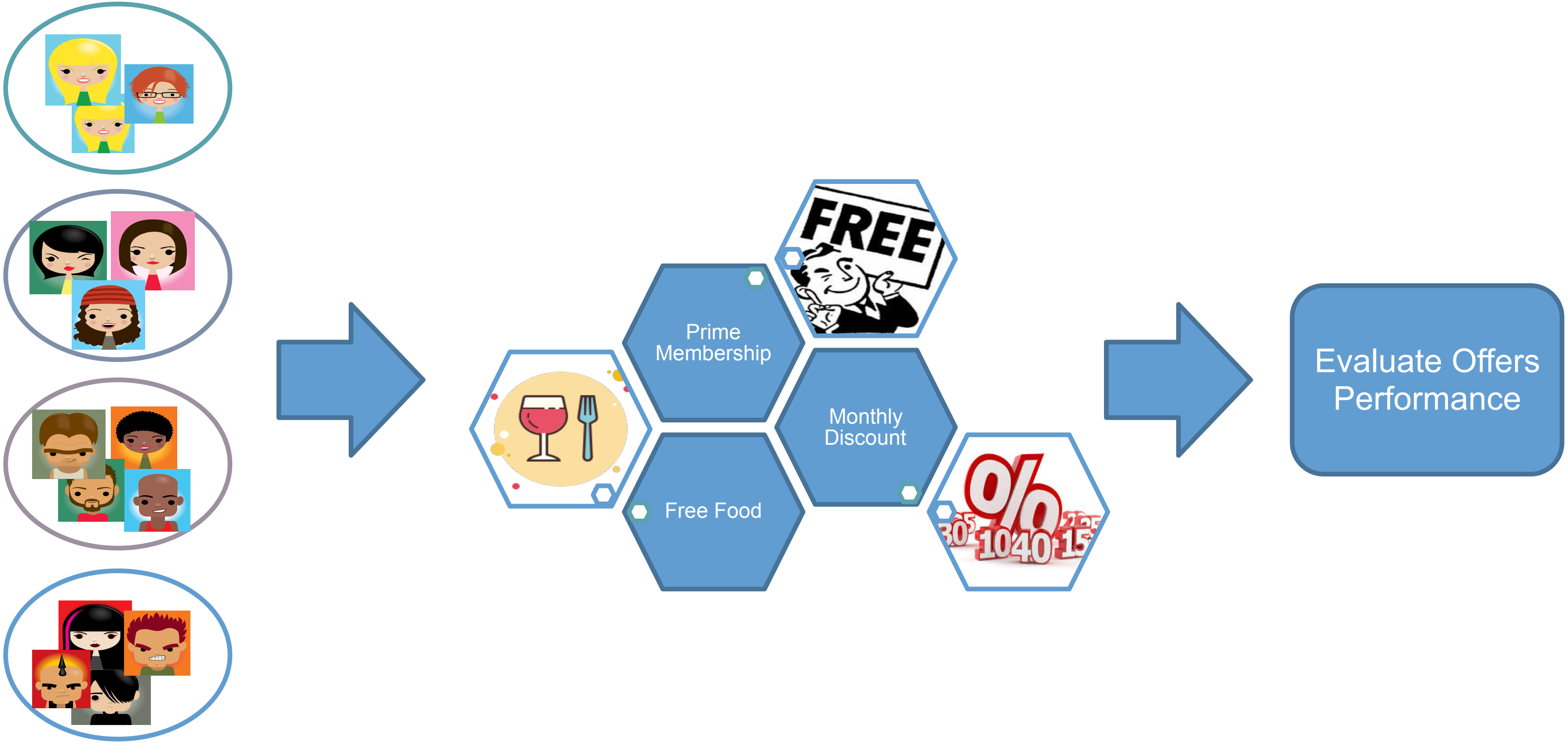
*Glady et al. (2009). Modeling churn using customer lifetime value.

Financial Evaluation of a Campaign

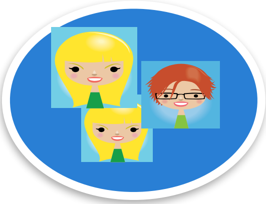
- Same offer may not apply to all customers (eg. Already is a prime subscriber)

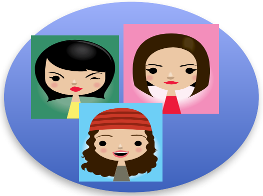
- An offer should be made such that it maximizes the probability of acceptance (γ)

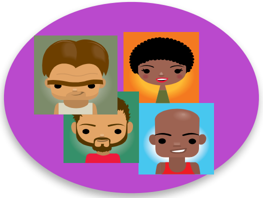
Offers Analysis

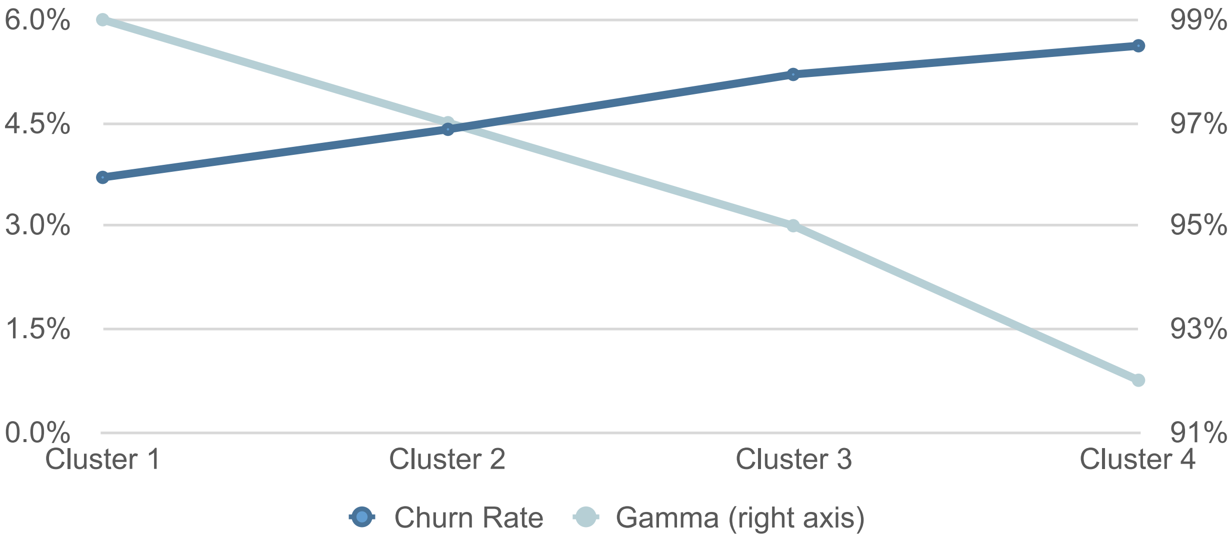


Offers Analysis









γ = Probability that a customer accepts the offer

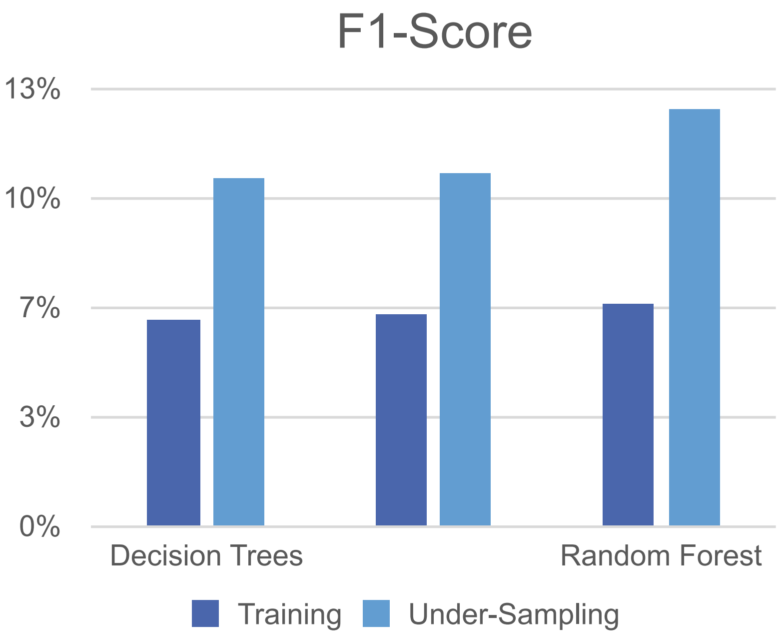
Predictive Modeling

Dataset	N	Churn	(Euros)
Total	9410	4.83%	580,884
Training	3758	5.05%	244,542
Validation	2824	4.77%	174,171
Testing	2825	4.42%	162,171
Under-Sampling	374	50.80%	244,542

Predictive Modeling

- Algorithms
    - Decision Trees
    - Logistic Regression
    - Random Forest

Predictive Modeling - Results



Predictive Modeling

- Sampling techniques helps to improve models’ predictive power however not necessarily the savings

- There is a need for methods that aim to increase savings

Agenda

- Churn modeling
- Evaluation Measures
- Offers
- Predictive modeling
- Cost-Sensitive Predictive Modeling
    - Cost Proportionate Sampling
    - Bayes Minimum Risk
    - CS – Decision Trees
- Conclusions

Cost Sensitive Predictive Modeling

- Traditional methods assume the same cost for different errors
- Not the case in Churn modeling
- Some cost-sensitive methods assume a constant cost difference between errors
- Example-Dependent Cost-Sensitive Predictive Modeling

Cost Sensitive Predictive Modeling

- Changing class distribution
  - Cost Proportionate Rejection Sampling
  - Cost Proportionate Over Sampling
- Direct Cost
  - Bayes Minimum Risk
- Modifying a learning algorithm
  - CS – Decision Tree

Cost Proportionate **Sampling**

- Cost Proportionate Over Sampling

Example		
1	0	1
2	1	10
3	0	2
4	1	20
5	0	1

Initial Dataset

(1,0,1)
(2,1,10)
(3,0,2)
(4,1,20)
(5,0,1)

Cost Proportionate Dataset

(1,0,1)
(2,1,1), (2,1,1), …, (2,1,1)
(3,0,2), (3,0,2)
(4,1,1), (4,1,1), (4,1,1), …, (4,1,1),
(4,1,1)
(5,0,1)

*Elkan, C. (2001). The Foundations of Cost-Sensitive Learning.

Cost Proportionate **Sampling**

- Cost Proportionate Rejection Sampling

Example			
1	0	1	0.05
2	1	10	0.5
3	0	2	0.1
4	1	20	1
5	0	1	0.05

Initial Dataset

(1,0,1)
(2,1,10)
(3,0,2)
(4,1,20)
(5,0,1)

Cost Proportionate Dataset

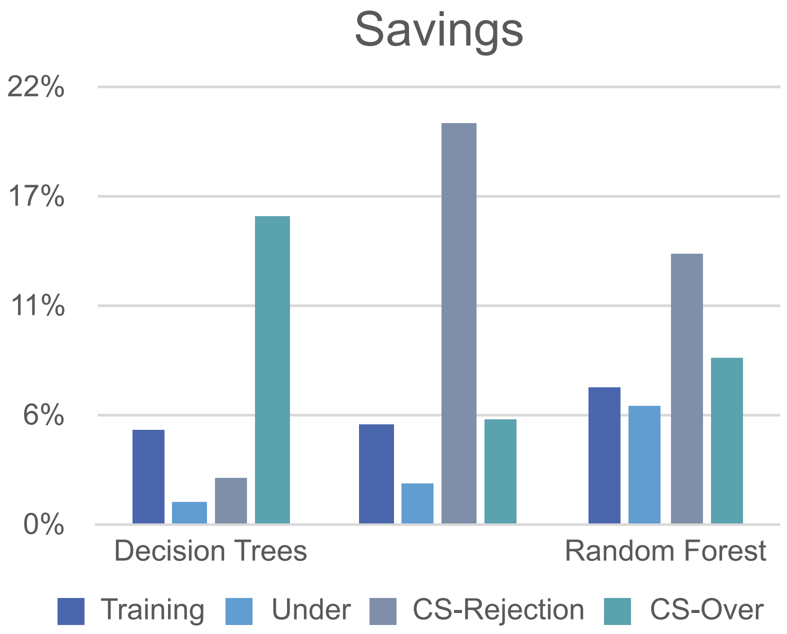
(2,1,1)
(4,1,1)
(4,1,1)
(5,0,1)

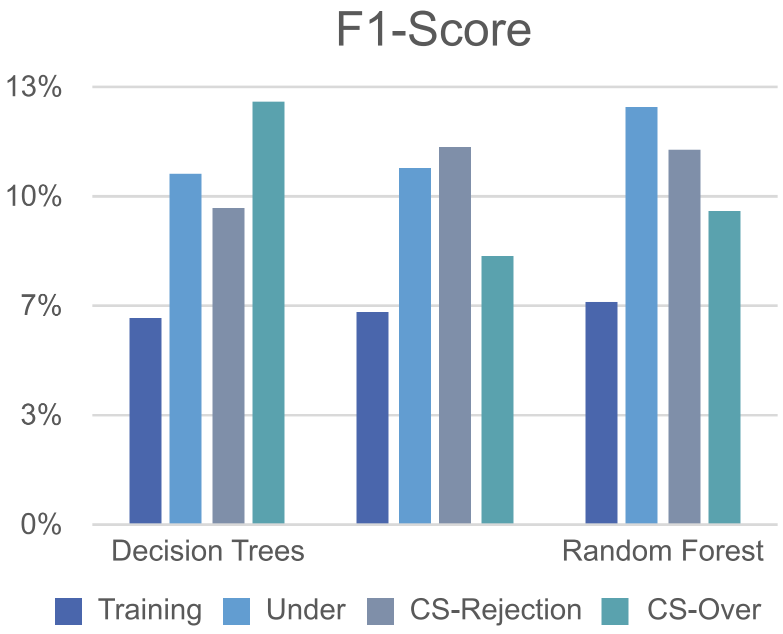
*Zadrozny et al. (2003). Cost-sensitive learning by cost-proportionate example weighting.

Cost Proportionate Sampling

Dataset	N	Churn	(Euros)
Total	9410	4.83%	580,884
Training	3758	5.05%	244,542
Validation	2824	4.77%	174,171
Testing	2825	4.42%	162,171
Under-Sampling	374	50.80%	244,542
CS – Rejection-Sampling	428	41.35%	231,428
CS – Over-Sampling	5767	31.24%	2,350,285

Cost Proportionate Sampling





Bayes Minimum Risk

- Decision model based on quantifying tradeoffs between various decisions using probabilities and the costs that accompany such decisions

- Risk of classification

$$R\left(c_i = 0 \mid x_i\right) = C_{TN_i}\left(1 - \hat{p}_i\right) + C_{FN_i} \bullet \hat{p}_i$$
$$R\left(c_i = 1 \mid x_i\right) = C_{FP_i}\left(1 - \hat{p}_i\right) + C_{TP_i} \bullet \hat{p}_i$$

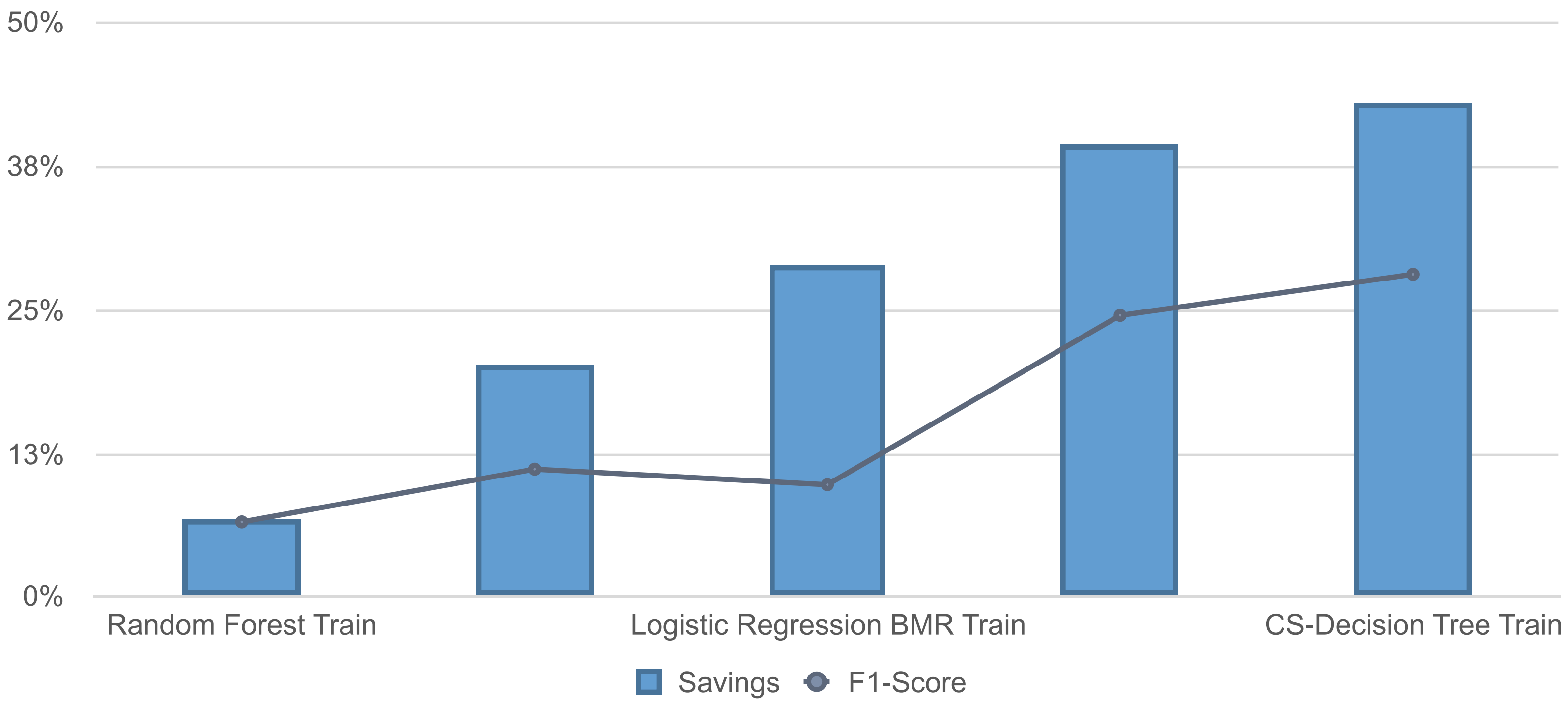
- Using the different risks the prediction is made based on the following condition:

$$c_i = \begin{cases} 0 & R\left(c_i = 0 \mid x_i\right) \leq R\left(c_i = 1 \mid x_i\right) \\ 1 & otherwise \end{cases}$$

Cost-Sensitive Decision Trees

- Decision trees
  - Classification model that iteratively creates binary decision rules $\left(x^j, l^j{}_m\right)$ that maximize certain criteria
  - Where $\left(x^j, l^j{}_m\right)$ refers to making a rule using feature j on value m

Comparison of Models



Conclusions

- Selecting models based on traditional statistics does not gives the best results measured by savings
- Incorporating the costs into the modeling helps to achieve higher savings

Correa Bahnsen *et al. Decision Analytics* (2015) 2:5
DOI 10.1186/s40165-015-0014-6

Decision Analytics
a SpringerOpen Journal

RESEARCH Open Access

A novel cost-sensitive framework for customer churn predictive modeling

Alejandro Correa Bahnsen*, Djamila Aouada and Björn Ottersten

*Correspondence:
alejandro.correa@uni.lu
Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg City, Luxembourg

Abstract
Customer churn predictive modeling deals with predicting the probability of a customer defecting using historical, behavioral and socio-economical information. This tool is of great benefit to subscription based companies allowing them to maximize the results of retention campaigns. The problem of churn predictive modeling has been widely studied by the data mining and machine learning communities. It is usually tackled by using classification algorithms in order to learn the different patterns of both the churners and non-churners. Nevertheless, current state-of-the-art

#EntregamosConAmor