POLITECNICO MILANO 1863 | DIPARTIMENTO DI INGEGNERIA GESTIONALE

Contractual settings

Churn prevention and survival analysis

Marketing Analytics

Customer Churn

Industry	Year(s)	Annual churn (%)
Internet service	2000-2002	21 - 63.2
Wireless Telephone	1999-2006	10 – 46
Satellite TV/radio	1999-2005	17 - 28
Financial services	1996-2001	6 - 30
Digital services	2005	20 - 46
Healthcare insurance	2004-2006	7.1 – 33.9
ADSL product	2006	6.1

Customer acquisition → Customer retention → Customer churn → Customer win-back

Customer Churn

Impact of churn

- Loss of revenue from defected customers
- Loss of opportunity to recover acquisition costs
- Loss of opportunity to cross-sell and to up-sell
- Negative WOM more likely than positive one
- New acquisition costs to replace lost customers

Customer Churn

Objectives of churn analytics:

- Why do customers churn?
- Which customers are in risk of churning?
- How can we retain the customers in risk of churning?

Customer Churn

Source of data

Total
customer base

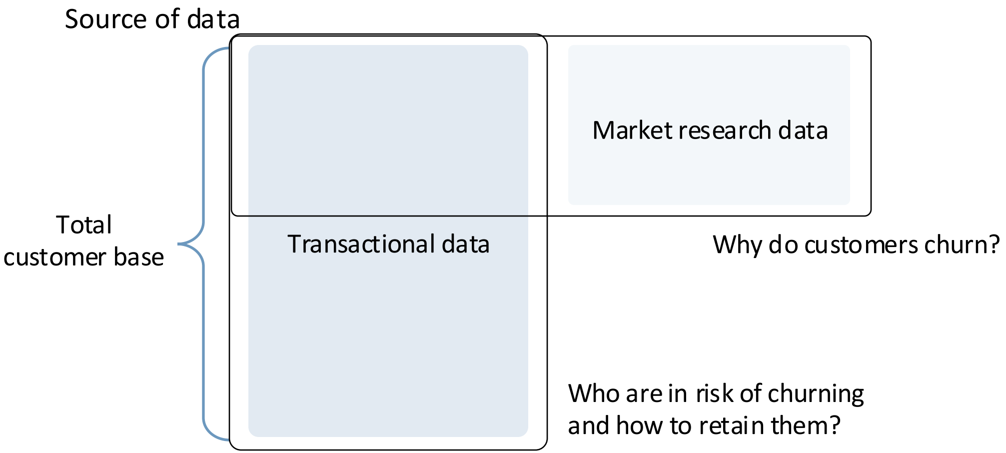
Transactional data

Market research data

- Satisfaction
- Perception
- Attitudes
- Needs
- ...

Customer acquisition → Customer retention → Customer churn → Customer win-back

Customer Churn



Customer Churn

Why do customers churn? Potential variables:

Customer relationship perceptions:

- Satisfaction (of the past)
- Commitment/trust (for the future)
- Contractual relationship (lead to apparent commitment)
- Price perception

Customer Churn

Why do customers churn? Potential variables:

Marketing actions:

- Direct marketing communication
- Mass marketing communication
- Loyalty and relationship programs
- Price tactics

Customer acquisition

Customer retention

Customer churn

Customer win-back

Customer Churn

Why do customers churn? Potential variables:

Moderating variables:

- Customer characteristics (demographics & psychographics)
- Environmental variables (e.g. competition)

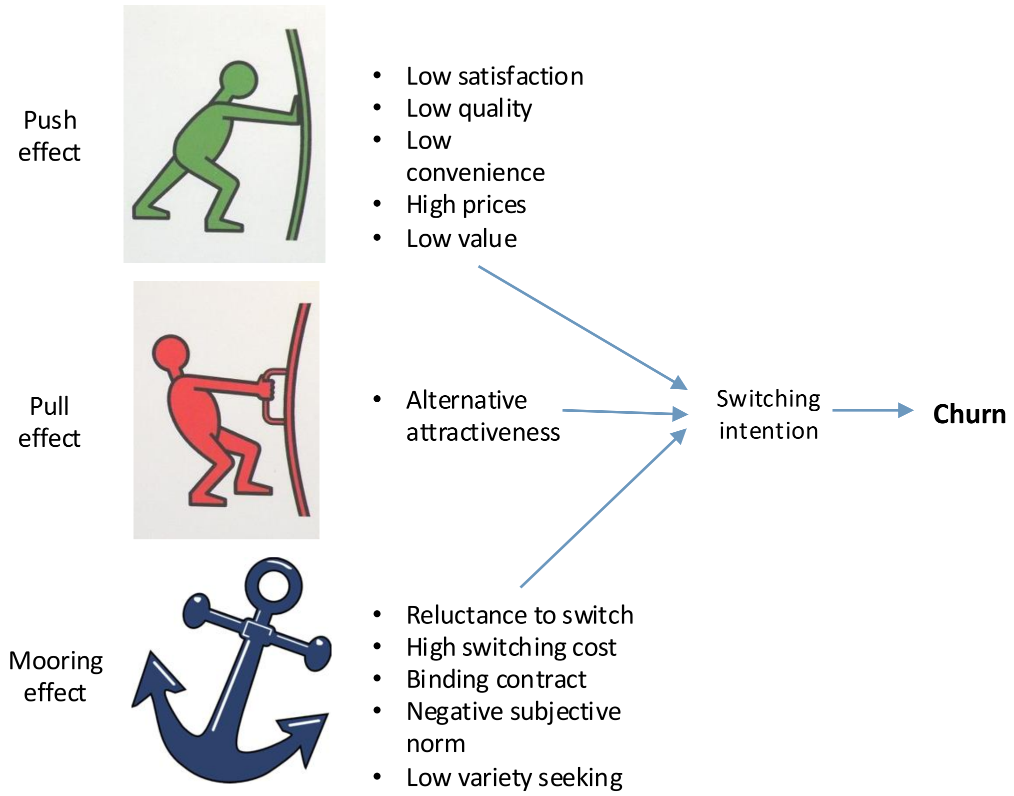
Customer Churn

Why do customers churn?

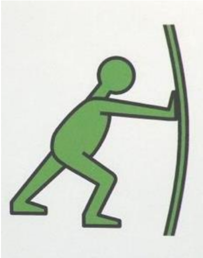
Market research data is not always available or updated. Proxies may be derived from transactional data. For example:

Proxies for satisfaction
- Complaints
- Cross-buying and up-buying
- Buying frequency trend
- Ticket size trend
- …

Proxies for commitment
- Subscription to loyalty program
- Saving personal/credit card info
- Frequency of visit/purchase
- Responding to marketing action
- …


Push effect
Low satisfaction
Low quality
Low convenience
High prices
Low value
Pull effect
Alternative attractiveness
Switching intention
Churn
Mooring effect
Reluctance to switch
High switching cost
Binding contract
Negative subjective norm
Low variety seeking

Push effect



- Low satisfaction
- Low quality
- Low convenience
- High prices
- Low value

Pull effect

- Alternative attractiveness

What could be the variables predicting churning among telecommunication customers?

Identify the source of the data, or the possible proxies.

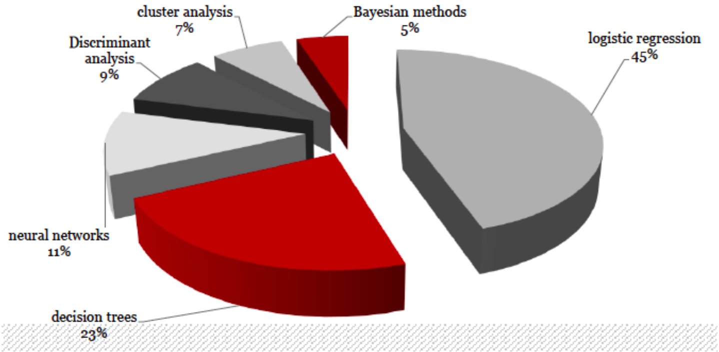
Mooring effect

- Reluctance to switch
- High switching cost
- Binding contract
- Negative subjective norm
- Low variety seeking

Customer Churn

Why do customers churn?

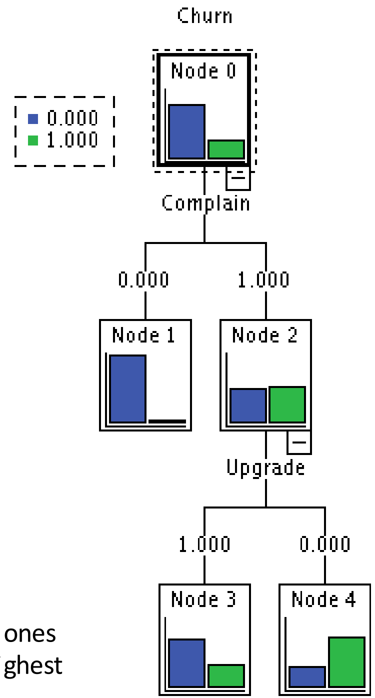
Statistical models



Neslin, et al., 2006

Decision Tree Analysis

Customer	Complain	Upgrade	Churn
1	0	1	0
2	0	0	0
3	0	0	1
4	1	0	1
5	1	1	0
6	0	1	0
7	1	0	1
8	0	0	0
9	0	1	0
10	1	1	0
...	...	...	...



Complain is the main trigger of churn.

Among those who have complained, the ones who have never had an upgrade are in highest risk of churning.

Customer Churn

Why do customers churn?

Identify variables and build models on calibration sample

Which customers are in risk of churning?

Test the model with validation sample, set up prediction protocol

How can we retain them?

Targeted action for customer retaining

Churn analysis

Logistic regression

Basic idea: customers can be split into two cateogries, in a given moment (T)

- **customers going to churn** (Y= 0)
- **customers not going to churn** (Y = 1)

Let x_i be an array of predictors

The likelihood of churn is modeled as a **conditional probability**, i.e.

$$\Pr(Y = 0 | x_i) = \frac{e^{-\beta_0 + \beta^T x_i}}{1 + e^{-\beta_0 + \beta^T x_i}}$$

Where β_0 represents the intercept and β is the vector of regression coefficients for x_i

Please refer to logistic regression notions for technicalities

Churn analysis

Genetic models

Basic idea: churn and no-churn are the consequence of a number of factors combined one with another, so no-churn (or churn) are the result of a «natural selection» of factors (predictors). The genetic model aims at identifying combination of predictors maximizing the probability of (no-)churn following a meta-heuristic model:

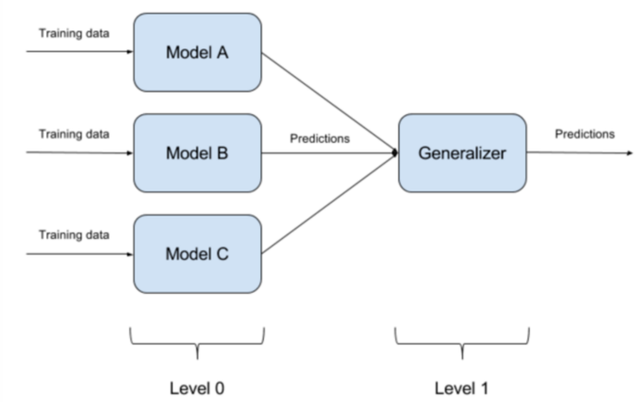
0. Let's start from a given population in which we know who churned and who did not
1. Random generation of a first set of solutions («chromosome», made of many components called «genes»)
2. Application of a fitness funtion (i.e. a metrics for understanding the predictive power of the actual churn/no-churn outcome)
3. Selection of the best solutions among the first chromosomes
4. *Crossover*, i.e., combination of chromosomes and genes
5. Identification of a population in the CB with the characteristics deriving from crossover
6. Repeat from step 2 on.

Possibility to introduce *mutations*, i.e. pseudo-random changes in the chromosomes to observe if the fitness improves

Churn analysis

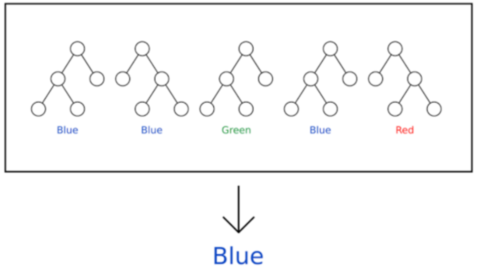
Random forest

Basic idea: **«A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.»**



Conceptually:

- The dataset is divided into a *training set* (where to «train» trees) and a *test set* where the rules are tested and proved for accuracy
- The algorithm selects random samples with replacement of the training set and develops trees on a limited set of features (bagging)
- These trees are processed with a parallel ensemble method to «compensate» errors



As per majority voting, the final result is 'Blue'.

Churn Analysis

XGBoost

Definition:
XGBoost (eXtreme Gradient Boosting) is a machine learning algorithm based on decision trees. It is designed for speed, efficiency, and performance, making it ideal for predictive analytics, including churn analysis. XGBoost excels in handling large datasets and complex relationships, offering high accuracy by optimizing gradient boosting through advanced regularization and computational efficiency.

Application in Churn Analysis:
In churn analysis, XGBoost models identify patterns and features that predict customer attrition. They provide actionable insights by ranking the factors most associated with churn, enabling targeted interventions to retain customers. Key advantages include:

- High Predictive Power: Captures non-linear relationships and interactions between variables.
- Feature Importance: Highlights the most significant factors driving churn, such as customer engagement, pricing, or service issues.
- Scalability: Efficiently handles large customer datasets typical in marketing analytics.

Churn Analysis

Key Parameters in XGBoost

1. Learning Rate (eta)

1. Controls the step size at each boosting iteration.
2. Smaller values lead to more conservative updates but require more iterations.
3. Typical range: 0.01 to 0.3.
4. **Impact on churn analysis:** A smaller learning rate may capture subtle patterns in churn-related data but increases computational time.

2. Number of Trees (n_estimators)

1. Defines the total number of trees (iterations).
2. Works in conjunction with the learning rate to balance model complexity and overfitting.
3. **Impact:** More trees generally improve performance but can lead to overfitting without proper regularization.

3. Max Depth (max_depth)

1. Sets the maximum depth of each decision tree.
2. Larger depth allows the model to capture more intricate patterns but may increase overfitting.
3. **Impact:** A well-tuned depth ensures the model captures churn drivers like customer behavior patterns without overfitting.

Churn Analysis

Key Parameters in XGBoost

4. Regularization Parameters

1. **L1 Regularization (alpha)**: Adds a penalty for the number of features used, encouraging sparsity.
2. **L2 Regularization (lambda)**: Penalizes large feature weights, reducing model complexity.
3. **Impact:** Reduces overfitting and ensures churn analysis generalizes well to unseen customers.

5. Subsample Ratio (subsample)

1. Controls the fraction of training data used for each tree.
2. Typical range: 0.5 to 1.0.
3. **Impact:** Improves model robustness by adding randomness, helping to avoid overfitting.

6. Colsample by Tree/Level (colsample_bytree, colsample_bylevel)

1. Specifies the fraction of features considered for each tree or tree level.
2. **Impact:** Reduces computation and enhances generalization in churn prediction.

7. Gamma (min_split_loss)

1. Minimum loss reduction required to make a split.
2. Larger values make the model more conservative, reducing the chance of overfitting.
3. **Impact:** Avoids unnecessary splits, making the model simpler and more interpretable for churn factors.

Churn Analysis

Key Metrics and Outputs in XGBoost

1.Gain

1. Measures the improvement in model performance due to a feature split in a tree.
2. **Impact:** Helps rank features by their importance in predicting churn (e.g., low customer engagement).

2.Cover

1. Represents the number of observations affected by a split.
2. **Impact:** Indicates how broadly a feature influences the model.

3.Weight

1. Reflects the number of times a feature is used in the trees.
2. **Impact:** Indicates a feature's relative importance in the model.

4.AUC-ROC (Area Under the Curve - Receiver Operating Characteristic)

1. Evaluates the model's ability to distinguish between churners and non-churners.
2. **Impact:** Higher AUC-ROC means better predictive power.

5.Log Loss

1. Measures the model's error for probabilistic outputs.
2. **Impact:** Lower log loss indicates better-calibrated churn predictions.

6.SHAP Values (Shapley Additive Explanations)

1. Quantifies each feature's contribution to individual predictions.
2. **Impact:** Explains why specific customers are flagged as likely to churn, supporting actionable interventions.

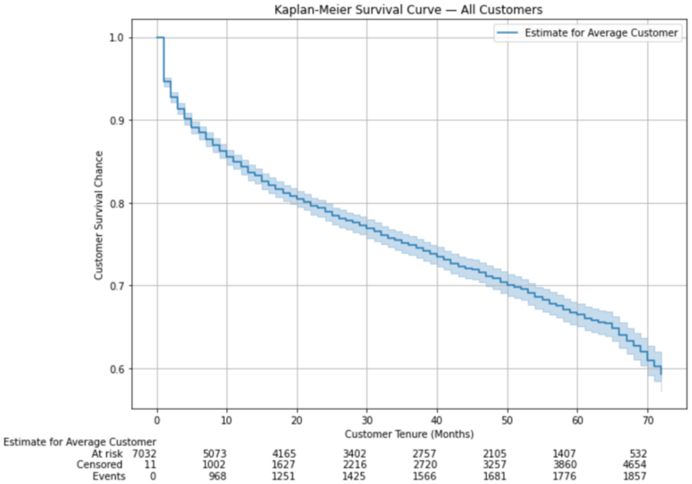
Churn analysis

Survival analysis

Basic idea: «churn» is conceptually equivalent to a death, and if we observe a sample of observation, we can calculate for any time interval t and for any lifetime T (continuous), the Survival Function:
S = P({T>t})

The survival function obviously decreases over time.

Survival analysis: an example

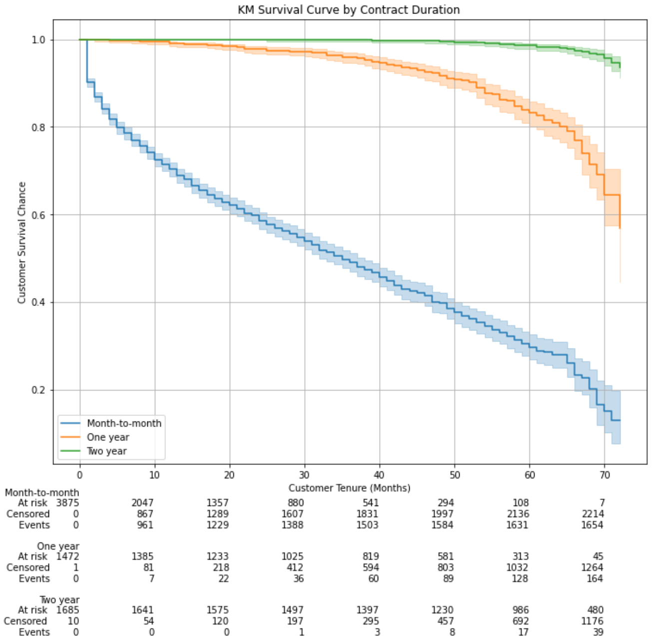


At-risk: number of customers with an observed tenure of more than the point in time.

Censored: number of customers with a tenure equal to or less than the point in time, which were not churned. For example, 3.860 customers had a tenure of 60 months or less However, they had not churned then

Events: number of customers with a tenure equal to or less than the point in time, which had churned by then. For example, 1,681 customers had a tenure of 50 months or less and had churned by then

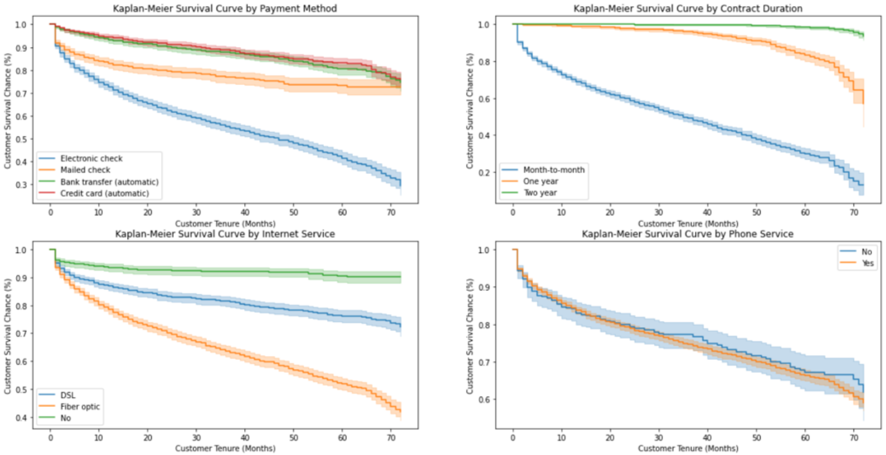
Survival analysis: an example



By comparing the survival functions with different values of the covariates (in this case, for instance, the contract length – i.e., monthly, yearly and 2-years), we can understand the relative impact of the covariates on tenure.
In order to assess this, we can calculate the **Hazard Function** (i.e., the probability that our event of interest occurs at a specific time) and the **Hazard ratio** (i.e., the ration between the hazard functions in correspondance to two different values of a covariate)

By calculating the hazard ratio for all the covariates we can understand which increase the most the probability to churn and the overall ability of the covariates to predict the event (i.e., churn)

Survival analysis an example



By comparing the survival functions we can understand which covariates better affect expected Lifetime. In this case:

- Automatic payment
- Yearly or bi-yearly contracts

There is some analysis to run to understand why DSL and Fiber drop survival

There is no evidence of a change in survival function in presence of a phone service

POLITECNICO MILANO 1863 | DIPARTIMENTO DI INGEGNERIA GESTIONALE

Churn prevention and survival analysis