Traditional Modelling Methods vs. Two-Stage Modelling – Marketing Implementation

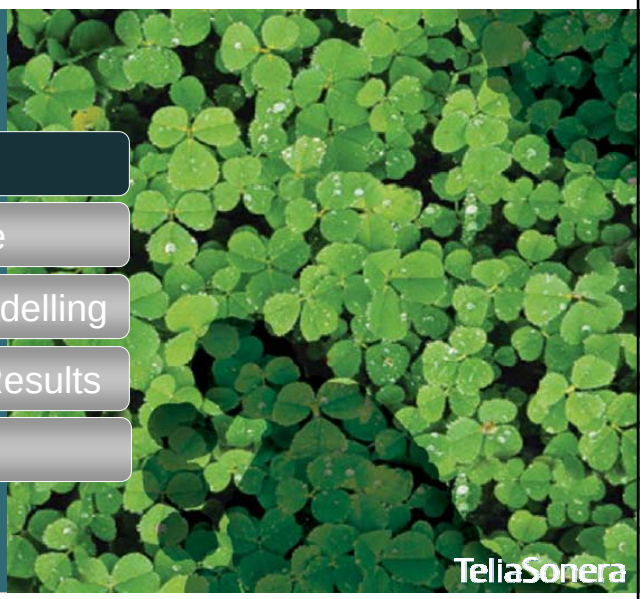
SAS Users' Spring Forum 2004 - Silja Serenade 21.-23.4.2004

Riku Mäkeläinen
TeliaSonera Finland / Consumer Marketing

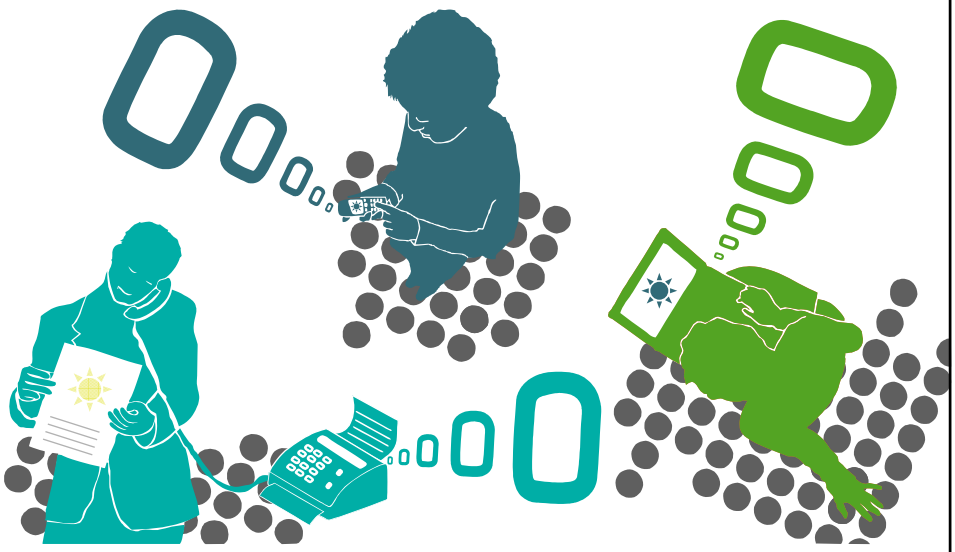
TeliaSonera

TeliaSonera

Contents



Introduction



Why Two-Stage Modelling

- to build better models
- another challenge
- better business

Id	Activation Probability	Predicted Revenue	Expected Revenue
12345	0,09	14,04 €	1,26 €
12346	0,71	1,17 €	0,83 €
12347	0,09	7,61 €	0,68 €
12348	0,71	7,61 €	5,40 €

Traditional activation models may overlook customers who are

- hard to activate - activation probability small
- but if activated yield high revenues - predicted revenue high

=> it makes sense to try to activate several hard to activate but potentially high value customers

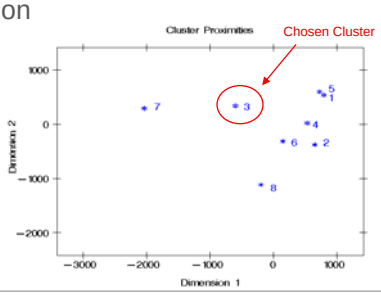
GOAL: Show how two-stage modelling increases revenues

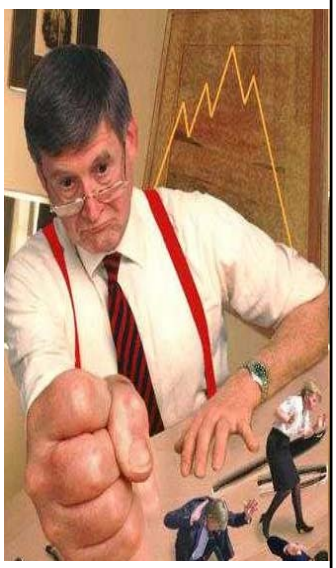
Business Case

- Business problem: Activate MMS usage profitably
- Solution: Targeted direct mail campaign

Actions:

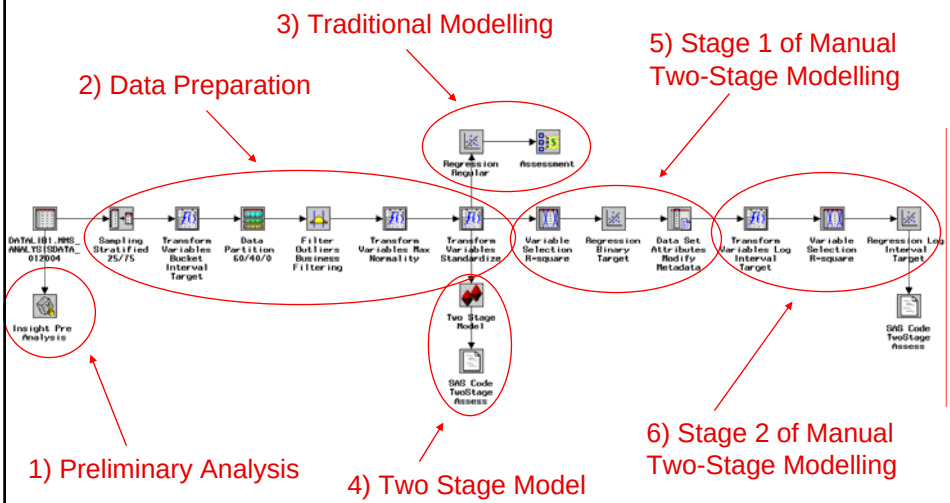
- Choose target segment / cluster
- Build predictive model
- Implement model
- Execute campaign
- Follow-up & Evaluation
- Automation





1. Introduction
2. Business Case
3. Two-Stage Modelling
4. Evaluation & Results
5. Summary

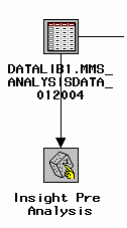
Overview of Analysis Chain



Preliminary Analysis

Analysis chain - 1/8

- **Input Data Source**: Input analysis data
  - target segment: 329 000 subscriptions with 1,77% target events
    - target: subscription that is actively using MMS – one-time users and random users excluded
  - profit matrix:
    - profit: successful contact yields 3,98 eur
    - cost: contact cost 0,70 eur/contact
- **Insight**: make preliminary data analyses

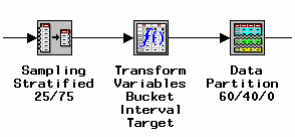


TeliaSonera

Data Preparation

Analysis chain - 2/8

- **Sampling**: Make stratified sample containing 25% target events and 75% non-target events
- **Transform Variables**: Bucket interval target to enhance data partition
- **Data Partition**: Split data to 60% training and 40% validation, test data not needed
  - use bucketed interval target to make sure both training and validation partitions have equal distribution on interval target



TeliaSonera

More Data Preparation

Analysis chain - 3/8

- **Filter Outliers**: Filter outliers based on statistics and business knowledge
  - knowing your data is mandatory!
- **Transform Variables**: Maximize normality and standardize variables to enhance modelling

- At this point we have dataset
  - we are familiar with
  - that has no outliers – we have the behaviour of the mass
  - which includes variables with 'nice' distributions

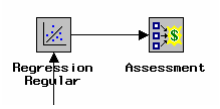
=> We are ready to do some serious modelling!

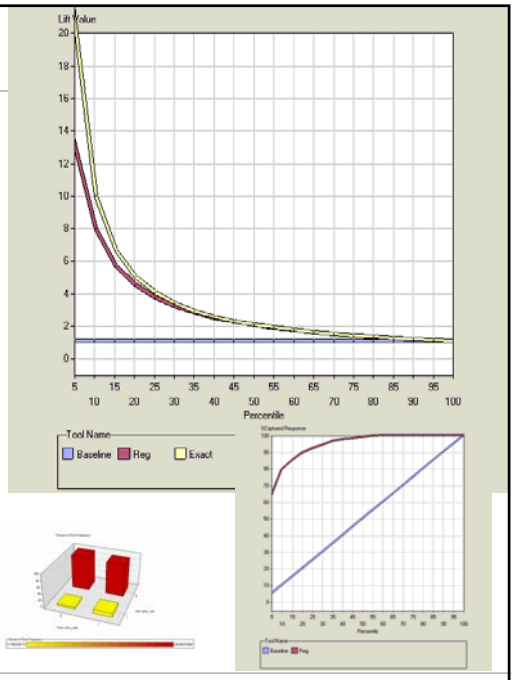
Regression Model

Analysis chain - 4/8

- **Regression:** Logistic regression
  - Stepwise selection method
  - Profit / Loss evaluation criteria
- **Assessment**: Results
  - Impressive Lift Value
  - Average profit 0,016386
  - Total profit 212,52

However, **we can improve!**

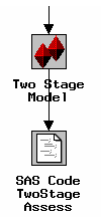




Two Stage Model Node

Analysis chain - 5/8

- **Two Stage Model**:
  - Stage 1 – activation probability: Decision Tree
  - Stage 2 – revenue from usage: Regression
- **SAS Code**: Model assessment
  - Adjust probability and profit calculations for separate sampling
  - Select subscriptions which satisfy
    activation_probability * predicted_revenue > contact_cost



- Results
  - Average profit 0,035734
  - Total profit 463,43

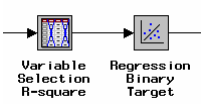
=> Increase of 118% in profits by using two stage modelling!

However, **we can improve!**

Manual Two Stage Modelling – Stage 1

Analysis chain - 6/8

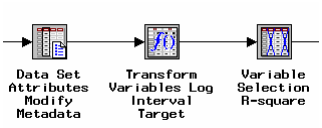
- **Variable Selection**: use R-square selection method to pre-select variables for modelling - override some rejections based on business knowledge
- **Regression**:
  - logistic regression with stepwise selection method
  - goal has shifted from making good decision to making an unbiased probability prediction -> use **validation error** criteria to evaluate model fit
  - enhance with significance levels, forced selection, optimisation criteria…
  - tip: Neural Networks or Tree-models may be useful if relation between inputs and target is non-linear
  - tip: avoid using minimize resource usage –option!



Preparation for Stage 2

Analysis chain - 7/8

- **Data Set Attributes**: modify metadata
  - model rejects variables not used, change status to input
  - coupling: add binary target predicted probability to list of inputs
- **Transform Variables**: log interval target
  - predicting negative usage amounts is not logical
  - by using log transformation we can correct error distribution of interval target and thus enhance model
- **Variable Selection**: use R-square selection method to pre select variables for modelling - override some selections based on business knowledge



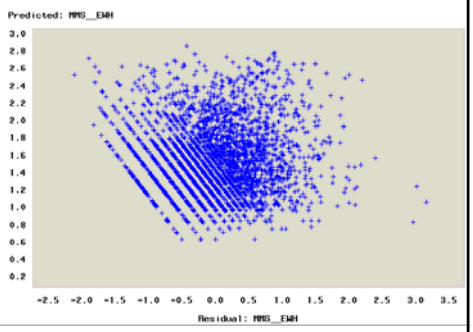
Manual Two Stage Modelling – Stage 2

Analysis chain – 8/8

- **Regression**:
  - linear regression with stepwise selection method
  - similar enhancements than with stage 1 logistic regression
    - log transformation helps correct error distribution
    - tip: try Poisson or Gamma distribution
- **SAS Code**: Assessment
  - Average profit 0,038320
  - Total profit 496,97

⇒ increase of 7,24% over Two Stage Model node

⇒ increase of 134% over regular regression



Summary of Modeling Results

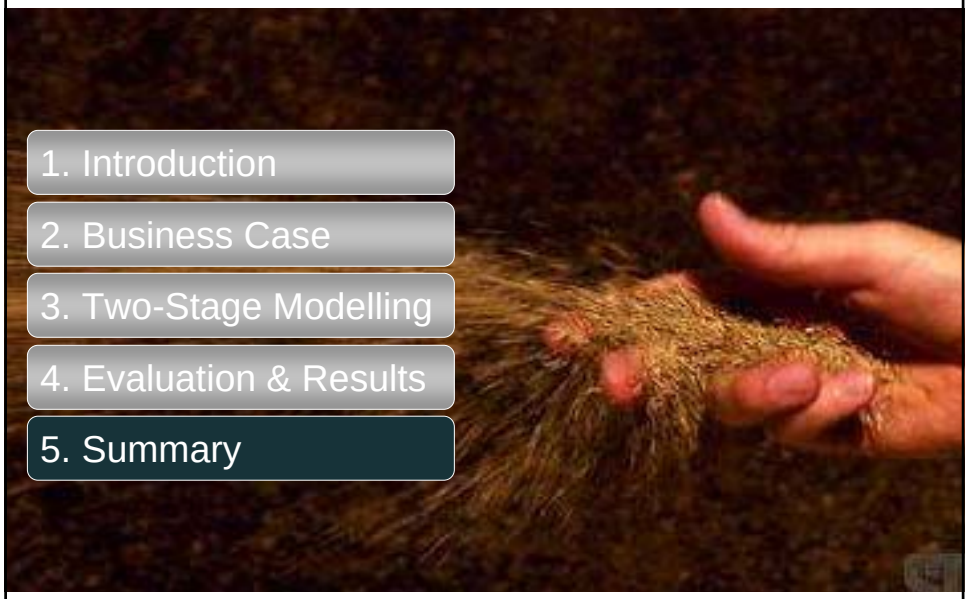
Method:	Total Revenue:	Average Revenue:
Regression	212,52	0,016386
Two Stage Model	463,43	0,035734
Manual Two Stage Model	496,97	0,038320

Note: revenues above represent only 4,68% of total due to sampling; magnitude of real profits are around 22 times higher

Evaluation

1. Two Stage Model node
   - easy and fast to use
   - limitations - reduced control over models
   - increases profits by 118%
2. Manual two-stage model
   - modelling takes time
   - extensive control over models and data
   - Increases profits by 134%

⇒ Does increase in profits justify time required for manual modelling?



Summary

- Definite improvement in target groups
- May require extra time
- Needs balancing between
  - results (+134% profits)
  - and requirements (time to model, understanding model)

The Most Important Thing

CUSTOMER